

RN [1] ?
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 8.2e-125;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTTHQSDVSYGVTVWELMTFGAKFYDGIAPAREIPDLLEKGERLPDPPICT 948
 DB 1 MALESILRRRFTTHQSDVSYGVTVWELMTFGAKFYDGIAPAREIPDLLEKGERLPDPPICT 60
 QY 949 IDVYIMVWKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEDELGPASPLDSTYRSI 1008
 DB 61 IDVYIMVWKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEDELGPASPLDSTYRSI 120
 QY 1009 LEDDDMGDLVDAEELVLPQGFPCPDPAAGGVVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDDMGDLVDAEELVLPQGFPCPDPAAGGVVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGY 1128
 DB 181 EAPRSPLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGY 240
 QY 1129 APLTCSQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSFGKGVVYKDVFAFG 1188
 DB 241 APLTCSQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSFGKGVVYKDVFAFG 300
 QY 1189 GAVENPEYLTPQGAAPOPHPHPAPSAFDNLYYWDQDPPEPRGAPSTFKGTPTAENPEY 1248
 DB 301 GAVENPEYLTPQGAAPOPHPHPAPSAFDNLYYWDQDPPEPRGAPSTFKGTPTAENPEY 360
 QY 1249 LGLDVVPV 1255
 DB 361 LGLDVVPV 367

RESULT 10

OSMYYVO PRELIMINARY; PRT; 412 AA.

ID OSMYYVO
 AC OSMYYVO;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PP3659
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318349; AAL55856.1; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FE3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.5e-121;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTTHQSDVSYGVTVWELMTFGAKFYDGIAPAREIPDLLEKGERLPDPPICT 948
 DB 1 MALESILRRRFTTHQSDVSYGVTVWELMTFGAKFYDGIAPAREIPDLLEKGERLPDPPICT 60
 QY 949 IDVYIMVWKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEDELGPASPLDSTYRSI 1008
 DB 61 IDVYIMVWKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEDELGPASPLDSTYRSI 120
 QY 1009 LEDDDMGDLVDAEELVLPQGFPCPDPAAGGVVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDDMGDLVDAEELVLPQGFPCPDPAAGGVVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGY 1128
 DB 181 EAPRSPLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGY 240
 QY 1129 APLTCSQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSFGKGVVYKDVFAFG 1188
 DB 241 APLTCSQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSFGKGVVYKDVFAFG 300
 QY 1189 GAVENPEYLTPQGAAPOPHPHPAPSAFDNLYYWDQDPPEPRGAPSTFKGTPTAENPEY 1248
 DB 301 GAVENPEYLTPQGAAPOPHPHPAPSAFDNLYYWDQDPPEPRGAPSTFKGTPTAENPEY 360
 QY 1221 YVWD-QDPPER-----GAPSTFKGTPTAEN 1245
 DB 361 YVWDQCEPBEQVRRSPDVSSGREGTSGAKRMEDPPTTSKGTCHARN 410

RESULT 11

OSMYYVO PRELIMINARY; PRT; 729 AA.

ID OSMYYVO
 AC OSMYYVO;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retrovirda; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynosccheck C., Jananson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSSP; P03322; 1A6S.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; Tyrosine-kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EF81D63 CRC64;

Query Match 23.6%; Score 1609; DB 15; Length 729;
 Best Local Similarity 52.5%; Pred. No. 2.1e-114;

	Matches	345,	Conservative	76;	Mismatches	126,	Indels	110;	Gaps	17;
QY	569	PONGSVTCRGPADCCACATKYPKPCVCARCBSGKYKPLDSTYPIKPKFDEBGAQQPPI	628							
Dd	141	PEETATPTKTP--DHCKCAHFIDGFHCVCACPGVLGENDTL-VKKYADANAVQLCHP	197							
QY	629	NCTHSCVDLDDEKCCEAQRASPLTISIYSAVY-GILLVVVLGVVFGLIRROOKRKRTM	687							
Dd	198	NCTRGGCKRGEGEC---NSKRTSISAGVYGGLCLLVVGIGIGIYLARR-HYRKRTL	253							
QY	688	RRLLETELVEEDLTSGAMPNQAOQRIKETELRKVKVLSGAFGNNTFSFWLRVRY	747							
Dd	254	RRLLORELVEEPLTSGCAPNOAHILRIKETEFKKYVSLGSFAFG---TVYKGIMIEG	309							
QY	748	SASHLEY-----IRENTPSKAKKEILIDEAYVWAGSPYSRLIGLCITSTOVLTMQMPY	803							
Dd	310	EKVYIPVAIKELREKTSPPANKELLIDEAYVASVDNRVRELGLCISTVOILTQMPLY	369							
QY	804	GCLLDHVRENGRSLGSODLLNMQIAKMSYLEDVLYHRDLAAENVLYKSPNHVKITD	863							
Dd	370	GCLLDVIRHKXONISQSXYLLMNVCQIAKMMVTEERLTYHRDLAARNVLYKTPOHVKITD	429							
QY	864	FGLIARLLDIDETFYHADGSKYPIKMMALESILTRRFTHOSDWVSQGYTWELMTGCAPY	923							
Dd	430	FGLIAKLGLADEKEYHAEGGVKPIKMMALESILHRIYHOSDWVSQGYTWELMTGSKPY	489							
QY	924	DGIPAEIEFDLLEKGERLPQPICITDIYVMIVMCMIIDSECFPRFEVLSESRMARDP	983							
Dd	490	DGIPAEISISYLEKERLPQPICITDIYVMIVMCMIIDDSRPKRRELIJAERSKVARDP	549							
QY	984	QRFVVIQ-NEDUGPASPLDTSTYRSLEDMDGULVAEEVLYPQGCFPPDPAPAGSM	1044							
Dd	550	PRYLVIQGBERMHLPSPTSKFYRLIMEEDMEDVIDADEVLPHQGF-----	598							
QY	1043	VHHRHSSSTRSGCGGLTLGLEPSEEARPSPL-----APESEGSPDFPDGLMGAAAG	1097							
Dd	599	-----NSPST-----SRIFLSSLATSINNSATNCID-----RN	627							
QY	1098	LQSLPTHDSPFLORYSEDETVLPSET--DGYYAPLTCSPOPEYVNPQDVRPOPSPREG	1155							
Dd	628	GQHPRFREDSPVQRYSSDPTGNLFLEESIDDGFL--PAPEVYNQ--TMPKPS----	675							
QY	1156	PLFAARPAGATLERATLTSPOKNGVXKVF-----APCGAVENPEYL	1197							
Dd	676	-----TAMVONQNIYNISLTAISKLPMDSRVONSHTAVDNPEYL	715							
RESULT 12										
ID	086714	PRELIMINARY;	PRT:	567	AA.					
AC	086714;									
DT	01-NOV-1996	(TREMBLrel). 01, Created)								
DT	01-NOV-1996	(TREMBLrel). 01, last sequence update)								
DT	01-MAR-2002	(TREMBLrel). 20, last annotation update)								
DE	V-erbB	protein (Fragment).								
CN	V-ERBB.									
OS	Avian	fous-associated virus type 1..								
OC	Viruses; Retroid	viruses; Retroviridae; Alpharetrovirus.								
OX	NCBI_TaxID=11950;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	MEDLINE=94203659; PubMed=8152791;									
RA	Vennstrom B., Raymscheck C., Janson L., Doederlein G., Imotek V.,									
RA	Johnson A., Beug H.;									
RT	"Retroviral capture of c-erbB proto-oncogene sequences: rapid									
RT	evolution of distinct viral genomes carrying mutant v-erbB genes with									
RT	different transforming capacities."									
RL	Oncogene 9:1307-1320(1994).									
DR	EMBL: S69372; AAC60727.1; -.									
DR	HSSP: P11362; IJGK.									
DR	InterPro: IPR000719; Euk_Pkinase.									
DR	InterPro: IPR001245; Tyr_Pkinase.									
DR	Pfam; PF00069; pkinase; I.									

DR	PRINTS; PR00109; TYRKINASE.	DR	ProdDom; PD000001; Euk_pkinase; 1.
DR <td>SMART; SM00219; TyrcK; 1.</td> <td>DR<td>PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.</td></td>	SMART; SM00219; TyrcK; 1.	DR <td>PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.</td>	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR <td>PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.</td> <td>DR<td>PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.</td></td>	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	DR <td>PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.</td>	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM <td>tyrosine-protein kinase.</td> <td>FT<td>NON TER</td></td>	tyrosine-protein kinase.	FT <td>NON TER</td>	NON TER
SC <td>SEQUENCE 567 AA; 63390 MW; G6D9CBA7ADF725E1 CRC64;</td> <td></td> <td></td>	SEQUENCE 567 AA; 63390 MW; G6D9CBA7ADF725E1 CRC64;		
Query Match	23.6%; Score 1607; DB 15; Length 567;		
Best Local Similarity	53.1%; Pred. No. 2e-114; Indels 110; Gaps 17;		
Matches 144;	Conservative 74; Mismatches 120;		
QY	578 GPEADQCACAHYADPFCFVARCSPGKPLDLYPIWKFPDEBAGCPPCINCHSCVDL	637	
DB	1 GP--DHCKKCAHFIDGPHCVACPAVLGENDTL-VKRYADANAVCQLCHPNCIRGCKGP	57	
QY	638 DDKGCPAQRASPLTSLYSANV-GILLVVLGVVFGVILIKRROOKIKYMRRLLOTEL	696	
DB	58 GLEGGP--NQSCKPSIAAGVVGGLCLLVVGILGIGYLRRL-HYKRTILRLLOREL	113	
QY	697 VEPLTPSGAMNOAKMRLKTELTKKVKVLGSGAFGFENFVSFWLVPKVASHLEY-	754	
DB	114 VEPLTPSGEAPNOAHLRLKTEFKKVKVLGSGAFG---TVYGMWIPBEKVKIPAI	169	
QY	755 --LRENTSPKANKETLDEAYVMAGVSGSPYVSRLLIGICTSTVOLVTQIMPYGCLLDHYRE	812	
DB	170 KELREATSPKANKETLDEAYMAVDNPRVCRLLIGICTSTVOLVTQIMPYGCLLDYRE	229	
QY	813 NRGRLGSDLLNMCQIAKGMSTYEDRLVLRDLAARVLYKSPNHYKITDGLARLLDI	872	
DB	230 HKDNIQSYLLNMCVQIAKGMNLYEERLVLRDLAARVLYKTHQVKITDGLAKLIGA	289	
QY	873 DETEYHAGCKVPIKMMALLESILRRFTHSDVMSYGVTVMLTFGAKPYDGIIPAREIF	932	
DB	290 DEKEVHAAGCKVPIKMMALLESILRRFTHSDVMSYGVTVMLTFGSKPYDGIIPAREIS	349	
QY	933 DLLEKGERLPQPPITTIIVYMTIMYKCMKIDSECPRPREFLYSEFSRMARDQRFVITQ-N	991	
DB	350 SYLKEGERLPQPPITTIIVYMTIMYKCMKIDSRKREFELLAEFSKMRADPRRYLVIGD	409	
QY	992 EDLGASPLDSTFTYSLLEDMDMGDLVAAEVLVPOQGFPCPDPAFGAGVNHRRSSS	1051	
DB	410 EEMHLPSPLDSTFTYTLMEEDMEDIVDAEYLVDHQFF-----NSPS	453	
QY	1052 TRSGGADLTLLGLEPSEEAAPSPL-----APSEGASDVPFGDLGMAKGLQSLPTHPD	1106	
DB	454 T-----SRTPLSLSLSSATSNNSATNCID-----RNGQGHPRVED	487	
QY	1107 SFLQYSEDPYPLPSET--DSYVAFLCSQPEPEYVNPQVAPROPSPREBGLPAAPRAG	1166	
DB	488 SFVQRRSSPFTNPLEESTIDGFL---PAPRYNQ--LMPKPKS-----	526	
QY	1165 APLERAKTISPGKNGVYKDV-----AFGAVENDEYL	1197	
DB	527 -----TAMVQNQIYNNISLTAISKLPMSRYQNSHSTANDNDEYL	566	
RESULT 13			
ID	Q64895		
AC	Q64895; PRELIMINARY; PRT; 962 AA.		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	GAG, V-erb-A, v-erb-B protein.		
GN	GAG, V-erb-A, v-erb-B.		
OS	Avian erythroblastosis virus.		
OS	Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.		
OX	NCBI_Taxid=11661;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RX MEDLINE=90206603; PubMed=1969616;
 RA Bruckner A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B".
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -.
 DR HSSP; P10828; ZNLF.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000536; Hormone_rec_lig.
 DR InterPro; IPR001723; Stehmann_receptor.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR001628; Znf_Casteroid.
 DR Pfam; PF00104; hormone_rec.1.
 DR Pfam; PF00069; pkinase.1.
 DR Pfam; PF00105; zf-C4.1.
 DR PRINTS; PR00398; STRDHOMOMER.
 DR PRINTS; PR00447; STROIDPINGER.
 DR ProDom; PD000001; Euk_pkinase.1.
 DR ProDom; PD000035; Znf_Casteroid.1.
 DR SMART; SM00430; HOL1.1.
 DR SMART; SM00219; Tyrc; 1.
 DR SMART; SM00399; ZNF_C4.1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR.1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108920 MW; 3C5AED791BE95CE CRC64;
 Query Match 22.7%; Score 1547.5; DB 15; Length 962;
 Best Local Similarity 49.7%; Pred. No. 1.6e-109;
 Matches 346; Conservative 74; Mismatches 149; Indels 127; Gaps 20;

541 VEEGRVGLPRE-YVNAR-HCLP-----CHPEQ 568
 354 IEKCOESTLAFEHYINRKNIHFMSKLMKVADLMGAYHASFLLMKVCEPTELS 413
 569 FONGSVTCFGEADQVCAAHYKDPFCVACPSGVKPDLSYPMKFPDEGACQPCPI 628
 414 PGE-----VGP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHP 465
 629 NCHTSQVLDLDDKCPAEGRASPLTISAVV-GILLVVVGVVFGILIKRQKIRKYM 687
 466 NCTRGCKGPGLEGCP--NSKTPSIAAGVGGLLCLVVGIGLGLRRLR-HIVKRTL 521
 688 RRLQETELVEPLTPSGAMPQAMRIKETELRKVKVLSGAFGNNFVSFWLVPKV 747
 522 RRLQERELVEPLTPSGAMPQAMRIKETELRKVKVLSGAFGNNFVSFWLVPKV 747
 748 SASHELV---LRENTSPKANKELIDEAYVMAGVSPYVRLIGICTSTVQLTQLMPEY 803
 578 EKVTIPVAIKELRATSPKANKELIDEAYVMASVDNPHVCRLLIGICTSTVQLTQLMPEY 637
 804 GCLLDHVENRGRGSOQLLNCMQIAKMSYLEDEVLVHRDLAARVLYKSNHVKIND 863
 638 GCLLDYIREKDNIGSQVLLNCMQIAKMSYLEDEVLVHRDLAARVLYKSNHVKIND 697
 864 FGLARLLDIDETEHADGKVPKIMMALESILRRRFTHSGDWSYGVYWEMLTFGAKPY 923
 698 FGLAKOLGADDEKEHAGKVPKIMMALESILRRRFTHSGDWSYGVYWEMLTFGAKPY 923
 924 DGIPAREIPDLLEKGERLPOPICTIVYMIYKCMWIDSECPREPREVSESRNARP 983
 758 DGIPAREIPDLLEKGERLPOPICTIVYMIYKCMWIDSECPREPREVSESRNARP 983
 984 QRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGDLVDAEYVLPQOGFCFDPAPGAGN 1042
 818 PRLVITQDERMHLPSPTDSKFRYTLMEEDMEDIVDAEYVLPQOGFCFDPAPGAGN 866

QY 1043 VHHRRSSSTRGGGDLTLGLEBSEBEPAPPLAPSEGASDVFDGLGGAAGLQSLP 1102
 DB 867 ---NSPST-----SSTPLLSLSASNS-----NSATKCIDRNG 896
 QY 1103 THDSPLORISEDETVLPSETDGYVAPLCSQPEFVYVNPDPVPPQPSREGPLPAAP 1162
 DB 897 GH-----PVREDFL-----PAPEVNO-LMPKPSYAVVQNOIYNY 932
 QY 1163 AGAT-LEBAKTLSPGKNVYKVDVAFGAVENPEYL 1197
 DB 933 ISLTAISLTPMDSRYQN-----SHTAVDNEEYL 961

RESULT 14

QY 085468 PRELIMINARY; PRT; 545 AA.
 AC 085468;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian Erythroblastosis virus (T834) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL; X06943; CA30024.1; -.
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; pkinase.1.
 DR ProDom; PD000001; Euk_pkinase.1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DCE8C8CAF8AF4 CRC64;

Query Match 22.6%; Score 1539; DB 15; Length 545;
 Best Local Similarity 52.7%; Pred. No. 3.1e-109;
 Matches 333; Conservative 71; Mismatches 128; Indels 100; Gaps 17;

578 GPEADQVCAAHYKDPFCVACPSGVKPDLSYPMKFPDEGACQPCINCHSVDL 637
 1 GP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHPCTRGCKGP 57
 638 DDKCPAEGRASPLTISAVV-GILLVVVGVVFGILIKRQKIRKYMRLQETEL 696
 58 GLEGCP--NSKTPSIAAGVGGLLCLVVGIGLGLRRLR-HIVKRTLRLQEREL 113
 697 VEPLTPSGAMPQAMRIKETELRKVKVLSGAFGNNFVSFWLVPKVSAHELV--- 754
 114 VEPLTPSGAMPQAMRIKETELRKVKVLSGAFGNNFVSFWLVPKVSAHELV--- 754
 755 --LRENTSPKANKELIDEAYVMAGVSPYVRLIGICTSTVQLTQLMPEYGLDHVE 812
 170 KELREATSPKANKELIDEAYVMASVDNPHVCRLLIGICTSTVQLTQLMPEYGLDYLRE 229
 813 NRGRLGSOQLLNCMQIAKMSYLEDEVLVHRDLAARVLYKSNHVKITDGLARLDI 872
 230 HKDNIGSQVLLNCMQIAKMSYLEDEVLVHRDLAARVLYKSNHVKITDGLARLDI 872
 873 DETEHADGKVPKIMMALESILRRRFTHSGDWSYGVYWEMLTFGAPYDGIAPAREIP 932
 290 DEKYHAGGKVPKIMMALESILRRRFTHSGDWSYGVYWEMLTFGAPYDGIAPAREIP 932

```

QY 933 DLKGERLLPQPICTIDVIMKCMWIDSECPREPRELVSERSWARDPQRFVIG-N 991
DB 350 SVLEKGERLPQPICTIDVIMKCMWIDSECPREPRELVSERSWARDPQRFVIG-N 409
QY 992 EDLGPASPLDSTFYRLSLDDMDGLVDAEYLYVPOQGFPCPDPAAGMVAHHRSSS 1051
DB 410 ERMHLPSPDTSKFYFTLMEEDMEDIVADLYVPHQOFF-----NSPS 453
QY 1052 TRSGGGLTLGLEPSEEPASPL-----ASSEAGSDVDPDGLGMAGCLOSLPHDP 1106
DB 454 T-----SRPFLSLISATSNNSANCIDRNG-----H-- 481
QY 1107 SPLQYSEDPVPLPSETDGVAPLTCSPQSEYVNOPDVPQPSPREGLPAPAPAGAT 1166
DB 482 -----PREDGFL-----PABEYVNO--LWPKKSTAMVQIQYINISLT 519
QY 1167 -LERAKTLPNGNVKGVKVFAGCAVENPEYL 1197
DB 520 AISKLPMSRYQN-----SHSTAVDNEPYL 544

```

RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.

AC Q9WVF5; 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor

isotorm 3).

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_Taxid=10090;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,

RA Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Mailhe N.J.;

RT "Alternative transcripts from the Human and Mouse EGFR Gene Encode

Carboxy-Terminal Truncated Receptors.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,

RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampand A.L.,

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

RA Mailhe N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

mouse alternative Egr transcripts encoding truncated receptor

isoforms.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

RA MEDLINE=21085660; PubMed=1127851;

RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,

RA Kuenli P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita A., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamitani M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzaletti U., Mombereis P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL; AF124513; AAD44149.1; .
DR EMBL; AF275366; AAG28047.1; .
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23681.1; .
DR EMBL; AK004883; BAB23641.1; .
DR EMBL; AK004911; BAB23662.1; .
DR MGI; 95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 22.1%, Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 1,3e-106;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

```

```

QY 11 LTLALPPGAA--STQVCTGDMKRLPASPETHLDMLRLHYOCQVQGNLETLVPTN 68
DB 14 LTLALCAAGALEKRYKQGSNRLTGLTFEDHFLSLQGMNNECEVLGNLETTYQRN 73
QY 69 ASLSFLQDIOEVQVYLIANQVPLQRLRYRGVQLFEDNYALAVLDNGPLNNTTP 128
DB 74 YDLSLTKTIOEVAGVLIANTVRIPLQIIRGNALVENVYALALSN----- 124
QY 129 VTGASPPGLRELQRLSTLEIKGVLQIQRNPOLCYOPTIIMKDI----FKKNQALTLI 184
DB 125 -YGVNRTGLRELPMRNLOEILIGAVRSNNPILCNMTTIQRPDIQVQVFNSSMSMDL--- 180
QY 185 DTRNSRACHPSPCKMGKGRSGESSDDQSLTRTVACGGA-RCKGPLPDCCHQCAAG 243
DB 181 -QSHSPSCPKDPCSPNCSWGSGEENCQKLTIKIICQCSHRRGRSPDCCNQCACG 239
QY 244 CGPRGSGCLACIHNHSGICELCPALVTYNTDPESSMNEGRYTFGASCYACPYNY 303
DB 240 CTGPRESCLVQCKQDEATCKDTCPLMLNPTTYQMDVNPBGKISFGATCVKCKERNY 299
QY 304 LSTDVGSCTLVCPHNGEVTAEQTCRCEKSKPCARVCYGLAQVYIKANSKFGITELE 363
DB 300 VVTDHGSQVRAQGPVYEV-BEDGIRKCKKDDGCRKVCNGIGIGERK-DTSLINATNIK 357
QY 364 -FAGCKKTFGSLAFIPESFDDPASNAPLOPBOLOVPELTLEETGLYISAPDLSPLD 422
DB 358 HFKYCTAISGDHLIPVAFKDSFTFRPPDPRELEILIKYKELTGFLLLIQWMDNNTDL 417
QY 423 SVFQNLQVIRGRIILHNGAYSLTLQGLGSMGLSLRELGSGLALIHNNHLCVHTVPW 482
DB 418 HAFENLEIRGRTKRGQFSLAVGLNITSLGLSLKEISGDVITSGNRLCYANTINW 477
QY 483 DGLFNPQALLHTANREDECVSEGLACHQLCARGHCWGPPTQCVNCSQPLRGQCEVE 542
DB 478 KKLFTLPQKTKIMNRRLEKCKKVNHNVCNPLCSSESCWGPEDPCVSCNVSGRGECVE 537
QY 543 EGRVYQGLPREVYVNRHLCPLHPCQPNVSGVTFGEADQVCVCAHYKDPPEVCNAPS 602
DB 538 KNLIEGPREVEVSECTIQHPELPPANNTITTGSPDICIICAHYIDGPHCVKTCIPA 597
QY 603 GVKPDLSTYMPIMKPPDEGAQCPPICTHSCVLDLDDKC 642
DB 598 GIMGENNTL-VWKYADANNVCHLCHANCTYCAAGPGLQGC 636

```

Search completed: July 22, 2003, 09:24:50
 Job time : 51.8246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.9339 Seconds

(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-710-730-14

Perfect score: 6813
Sequence: 1 MELALCRWGLLALPPGA.....TFKQTPAENPEYLGADVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*
1: /SID27/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID27/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID27/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID27/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID27/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID27/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID27/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID27/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID27/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID27/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID27/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID27/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID27/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID27/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID27/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID27/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID27/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID27/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID27/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID27/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID27/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID27/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID27/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	97.0	1255	21	AA92620
2	6608	97.0	1255	22	AAE12130
3	6608	97.0	1255	22	AAE12130
4	6608	97.0	1255	22	AAE12130
5	6602	96.9	1255	23	AAU74545
6	6602	96.9	1255	17	AAW01111
7	6602	96.9	1255	21	AAW01111
8	6602	96.9	1255	21	AAW01111
9	6602	96.9	1255	22	AAW01111
10	6602	96.9	1255	22	AAW01111

11	6602	96.9	1255	23	AAE24067	Human Her-2 protei
12	6602	96.9	1255	23	AAE20479	Human Her-2/neu pr
13	6602	96.9	1255	23	AAE11143	Human Her-2/neu on
14	6602	96.9	1255	23	AAU77114	Human Her-2/neu po
15	6559	96.3	1433	14	AAU95658	Sequence of c-erbB
16	6438	94.5	1223	23	AAU98923	Human breast cance
17	6285	92.3	1200	21	AAE21208	Human Her-2/neu pr
18	5815.5	85.4	1256	21	AAE21199	Rat Her-2/neu prot
19	5815.5	85.4	1256	23	AAE11144	Rat Her-2/neu onco
20	5788.5	85.0	1256	21	AAE11206	Mouse Her-2/neu pr
21	5788.5	85.0	1256	22	AAE20443	Mouse Her-2/neu pr
22	5788.5	85.0	1256	22	AAE11151	Mouse Her-2/neu on
23	4814	70.7	919	21	AAE21203	Human Her-2/neu fu
24	4814	70.7	919	23	AAE11158	Human Her-2/neu ex
25	4064.5	59.7	920	23	AAE11152	Mouse Her-2/neu ex
26	4064.5	59.7	926	23	AAE11153	Mouse Her-2/neu ex
27	3698	54.3	712	21	AAE21204	Human Her-2/neu fu
28	3698	54.3	712	23	AAE11149	Human Her-2/neu ex
29	3552	52.1	782	18	AAE19764	Her-2/neu extracel
30	3550	52.1	653	21	AAE21200	Extracellular HER-
31	3550	52.1	653	23	AAE11145	Human Her-2/neu on
32	3512	51.5	645	22	AAE20408	Human ErbB2 oncopr
33	3512	51.5	645	22	AAE11593	Human ErbB2 extrac
34	3447	50.6	951	21	AAE44993	DC86CFV-erbB2EC fu
35	3344	49.1	624	11	AAE08222	Extracellular port
36	3053.5	44.8	654	21	AAE21205	Rat Her-2/neu prot
37	3053.5	44.8	654	23	AAE11150	Rat Her-2/neu onco
38	3035	44.5	1210	21	AAE19259	Amino acid sequenc
39	3035	44.5	1210	21	AAE19259	Human EGF receptor
40	3035	44.5	1210	23	AAE203019	Human Her-1 protei
41	3035	44.5	1210	23	AAE07658	Human epidermal gr
42	3033	44.5	1210	22	AAE68420	Amino acid sequenc
43	2994	43.9	1210	23	AAE51768	Human epidermal gr
44	2958	43.4	583	23	AAE20443	Human protein for
45	2958	43.4	587	23	AAE20461	Human protein for

ALIGNMENTS

RESULT 1
AA92620
ID AA92620 standard; Protein; 1255 AA.
AC AA92620;
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
FH Key
FT Domain
FT
FT
FT Region
FT
FT Region
FT
FT Region
FT
FT Region
FT
FT Region
FT
FT Region
FT
FT Region
FT
FT Domain

Location/Qualifiers
1..173
/label= N-terminal
/note= "mature polypeptide"
5..25
/label= insertion region
/notes= "suitable for foreign epitope insertion"
59..73
/label= insertion region
/label= "suitable for foreign epitope insertion"
103..117
/label= insertion region
/label= "suitable for foreign epitope insertion"
149..163
/label= insertion region
/label= "suitable for foreign epitope insertion"
174..323

FT /label= Cysteine_rich_domain
 FT Region 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 FT /label= ligand_binding_domain
 FT Region 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 FT /label= Cysteine_rich_domain
 FT Region 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 FT /label= Transmembrane_domain
 FT Region 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 FT /label= Tyrosine_kinase_domain
 FT Region 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 FT /label= C-terminal_domain
 PN WO200020027-A2.
 PD 13-APR-2000.
 XX
 PF 05-OCT-1999; 99WO-DK00525.
 XX
 PR 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX
 PA (MEB1-) M & E BIOTECH AS.
 PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI, 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198; 220pp; English.
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2), and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animal's immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SO Sequence 1255 AA;
 Query Match 97.0%; Score 6608; DB 21; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;
 QY 1 MELALCGWGLLLALPPGASTQVCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
 DB 1 MELALCGWGLLLALPPGASTQVCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
 QY 61 ELTYLPTNASLSPFLODIEQVGYVLIANQVRQVPLQRLRIVRGTQLFEDNVALAVLDNG 120
 DB 61 ELTYLPTNASLSPFLODIEQVGYVLIANQVRQVPLQRLRIVRGTQLFEDNVALAVLDNG 120
 QY 121 DPLNNTVTGASRGRLRELRLTELKGVGVLIQNPOLCVQDITLMDIFPKXNQLA 180
 DB 121 DPLNNTVTGASRGRLRELRLTELKGVGVLIQNPOLCVQDITLMDIFPKXNQLA 180
 QY 181 LTLIDTNRSPRACHPCSPCKSGRCWGSSEDCOSLRTTVAGAGCARCKGPLPTDCHEQC 240
 DB 181 LTLIDTNRSPRACHPCSPCKSGRCWGSSEDCOSLRTTVAGAGCARCKGPLPTDCHEQC 240
 QY 241 AAGCTGPRHSDCLACLAHNSGICELHCPALVYNTPTFSPMPREGRYTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLACLAHNSGICELHCPALVYNTPTFSPMPREGRYTFGASCVTACP 300
 QY 301 VYVLTSDVSGCTLVCPFLNQEVTAEADGTORCEKSKCARCYGLQVYIKANSKFYGIT 360
 DB 301 VYVLTSDVSGCTLVCPFLNQEVTAEADGTORCEKSKCARCYGLQVYIKANSKFYGIT 360
 QY 361 ELPEAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFETLEITGYLYISAMPDPLP 420
 DB 361 IOEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFETLEITGYLYISAMPDPLP 420
 QY 421 DLVSFQNLQVIRGRILNHGAYSLTLOGIGISWGLRSLREISGLAIHHNTHLCPVHTV 480
 DB 421 DLVSFQNLQVIRGRILNHGAYSLTLOGIGISWGLRSLREISGLAIHHNTHLCPVHTV 480
 QY 481 PWDQLFRNPHQALHTANRPEDECVEGSLACHQLCARHGWCPPTCVNCSQFLRQEC 540
 DB 481 PWDQLFRNPHQALHTANRPEDECVEGSLACHQLCARHGWCPPTCVNCSQFLRQEC 540
 QY 541 VEECRVLOGLPREVYNARHCLPHECQPPQNGSVTCGPEADQCVAAAHKDPFCVARC 600
 DB 541 VEECRVLOGLPREVYNARHCLPHECQPPQNGSVTCGPEADQCVAAAHKDPFCVARC 600
 QY 601 PGVVPDLSYMPIMWFPDEGACOPCPINTCHSCVDLDDKCPAEOPASPLTISVAVVG 660
 DB 601 PGVVPDLSYMPIMWFPDEGACOPCPINTCHSCVDLDDKCPAEOPASPLTISVAVVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKRYMARLLQETELVEPLTPSAMNQOMRLKETEL 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKRYMARLLQETELVEPLTPSAMNQOMRLKETEL 720
 QY 721 RKVKVLGSGAFGNFTVSPWLRLVP-----KVASHEVLRENTSPANKELIDEAYVMA 775
 DB 721 RKVKVLGSGAG-----TYKGIWLPDGENVAPIVA-IKVLRENTSPANKELIDEAYVMA 775

QY 776 GVSGPYVSRLLGICLTSTVQVLTQMPYQGLDHYENRGRGLGSDLLNMQIAKMSY 835
 DB 776 GVSGPYVSRLLGICLTSTVQVLTQMPYQGLDHYENRGRGLGSDLLNMQIAKMSY 835
 QY 836 LEDVRLVHRDLAARNLVKSPNHVKITDPGLARLIDIDETEHADGKVPIKMALESIL 895
 DB 836 LEDVRLVHRDLAARNLVKSPNHVKITDPGLARLIDIDETEHADGKVPIKMALESIL 895
 QY 896 RRRFTHQSDVMSVGTVMELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYIM 955
 DB 896 RRRFTHQSDVMSVGTVMELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYIM 955
 QY 956 VKCMTIDSECRPFRELVSFESRMARDPQRFVYIIONEDLGASPIIDSTFYRSLLEDMDG 1015
 DB 956 VKCMTIDSECRPFRELVSFESRMARDPQRFVYIIONEDLGASPIIDSTFYRSLLEDMDG 1015
 QY 1016 DLVDAEEYLVPOGFCFCDPAPOAGGMVHRHRSSTSGGSDLTGLGEPEEAPSPSL 1075
 DB 1016 DLVDAEEYLVPOGFCFCDPAPOAGGMVHRHRSSTSGGSDLTGLGEPEEAPSPSL 1075
 QY 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPPLPSETDGYVAPLTCSP 1135
 DB 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPPLPSETDGYVAPLTCSP 1135
 QY 1136 QPEYVNPQVPRPQPSRREGSLPAARPAATLERAKTISPGKNGVYKQVFAFGAVENPE 1195
 DB 1136 QPEYVNPQVPRPQPSRREGSLPAARPAATLERAKTISPGKNGVYKQVFAFGAVENPE 1195
 QY 1196 YLTPGGAAPQPHPPAFSPAFDNLVYWDQDPERGAPPSTFKGPTAENBEYLGLDVYV 1255
 DB 1196 YLTPGGAAPQPHPPAFSPAFDNLVYWDQDPERGAPPSTFKGPTAENBEYLGLDVYV 1255

RESULT 2
 AAE12130
 ID AAE12130 standard; Protein; 1255 AA.
 XX
 AC AAE12130;
 DT 18-DEC-2001 (first entry)
 XX
 DE Human tyrosine kinase-type receptor, HER-2.
 XX
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 XX
 PN WO200168677-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 16-MAR-2001; 2001MO-US40328.
 XX
 PR 16-MAR-2000; 2000US-0527487.
 XX
 PA (GEN2) GENZYME CORP.
 XX
 PI Nicolette CA;
 XX
 DR WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX
 PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX
 PS Claim 4: Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterized by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridization probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 CC
 SQ Sequence 1255 AA;
 XX
 Query Match 97.0%; Score 6608; DB 22; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;
 QY 1 MELAIQRMGLLALLPQGAATQVCTGTMTKLRLPASPETHLMDLRLYGGCCVQGNL 60
 DB 1 MELAIQRMGLLALLPQGAATQVCTGTMTKLRLPASPETHLMDLRLYGGCCVQGNL 60
 QY 61 ELTYLPTNALSFLDIOEYQGVYLIANNOYQVLORLIVRGTLQFEDNYALAVLDNG 120
 DB 61 ELTYLPTNALSFLDIOEYQGVYLIANNOYQVLORLIVRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVGASPGGLRELQRLSTELIKGGVLIQRNPOLCYQDTILMKDI FHKNNQLA 180
 DB 121 DPLNNTPTVGASPGGLRELQRLSTELIKGGVLIQRNPOLCYQDTILMKDI FHKNNQLA 180
 QY 181 LTLIDNRBRACHPGSPMKSGRCMGESSESDQSLTRTVCAAGCARKCGPLPTDCCHQC 240
 DB 181 LTLIDNRBRACHPGSPMKSGRCMGESSESDQSLTRTVCAAGCARKCGPLPTDCCHQC 240
 QY 241 AAGCTGPKHSDCLACILFENHSGICLHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACILFENHSGICLHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300
 QY 301 YNVLSTDVGSCTLVCPRLHNOEYTAEDGTQRCGSKPCARVCYGLGMOYIYANSEFIQIT 360
 DB 301 YNVLSTDVGSCTLVCPRLHNOEYTAEDGTQRCGSKPCARVCYGLGMOYIYANSEFIQIT 360
 QY 361 ELFEFAGCKKI FGSILAFIPSPGDPASNTAPLQPRQLQVFETLEITGYLIYSAMPDSL 420
 DB 361 ELFEFAGCKKI FGSILAFIPSPGDPASNTAPLQPRQLQVFETLEITGYLIYSAMPDSL 420
 QY 421 DLSVFQNLQVIRGRIIHNGAYSLLTQGLGISWLGRLSRLGSGIALIHNHTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRIIHNGAYSLLTQGLGISWLGRLSRLGSGIALIHNHTHLCFVHTV 480
 QY 481 PMDQLFRNHQALLHTANRPEDECVGEGLAGHQLCARHCGPRTQCVNSQPIRGQEC 540
 DB 481 PMDQLFRNHQALLHTANRPEDECVGEGLAGHQLCARHCGPRTQCVNSQPIRGQEC 540
 QY 541 VEECRVLQGLPREYVNAARCLPCHHECOFQNSVYCFPEADQCAACHYKDPFCVAVC 600
 DB 541 VEECRVLQGLPREYVNAARCLPCHHECOFQNSVYCFPEADQCAACHYKDPFCVAVC 600
 QY 601 PSGVKKDLSYMPIMKFPDEEGACQCPINCHSCVDLDDKGPAPQASPLTISVAVG 660
 DB 601 PSGVKKDLSYMPIMKFPDEEGACQCPINCHSCVDLDDKGPAPQASPLTISVAVG 660
 QY 661 ILVVVVLGVVFGILLKRRQOKIRKTYMRLLOETELVPLTPSGAMPQAOARILKETEL 720
 DB 661 ILVVVVLGVVFGILLKRRQOKIRKTYMRLLOETELVPLTPSGAMPQAOARILKETEL 720

Qy	721	RKRYKVGSGAFGNNNTVSFWLRLVP-----KVSASHLEVALENTSPRANKHEILDEAVYMA	775
Db	721	RKYKVVGSGAFG----TVYKGIMIPDGENVKLPVA-IVKYLEENTSPRANKHEILDEAVYMA	775
Qy	776	GVGSPYVSRLLGICLTSTVQLVTQLMPYQCLLDHRENRGSLGSODLLNMCQIAKMSY	835
Db	776	GVGSPYVSRLLGICLTSTVQLVTQLMPYQCLLDHRENRGRLGSODLLNMCQIAKMSY	835
Qy	836	LEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVP1KMALESIL	895
Db	836	LEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVP1KMALESIL	895
Qy	896	RRRPTHSDVWSYGVTVWELMTFGAKPYGIPAREIPDLLEKGERLPQPICTIDVYIM	955
Db	896	RRRPTHSDVWSYGVTVWELMTFGAKPYGIPAREIPDLLEKGERLPQPICTIDVYIM	955
Qy	956	VKCMIMIDSECRPFRELVSFESFMADPQRFVYVIONEDLGPASPLDSTIFYSHLLEDDMG	1011
Db	956	VKCMIMIDSECRPFRELVSFESFMADPQRFVYVIONEDLGPASPLDSTIFYSHLLEDDMG	1011
Qy	1016	DLVVAEETVLPQGFPCPPAPAGAGMHHRRSSSTRSGGSDLLTGLSPSEEEAPRSPL	1071
Db	1016	DLVVAEETVLPQGFPCPPAPAGAGMHHRRSSSTRSGGSDLLTGLSPSEEEAPRSPL	1071
Qy	1076	APSEGASDVFDGDLGKGAKGHQS.LPTHDPSPLORYSEDPVPLPSEITDGYVALTCSP	1131
Db	1076	APSEGASDVFDGDLGKGAKGHQS.LPTHDPSPLORYSEDPVPLPSEITDGYVALTCSP	1131
Qy	1136	QPEYVNPDPVRPOPSPSPREGPLLEAAPPAGATLEBRAKTLSPGKNGVYKQVFAFGAVENPE	1191
Db	1136	QPEYVNPDPVRPOPSPSPREGPLLEAAPPAGATLEBRAKTLSPGKNGVYKQVFAFGAVENPE	1191
Qy	1196	YLTQGGGAQOPHPFPSPAFPNLTVYMODPPEERGAPSTFKGTPTAENPEYLGLOVPV	1251
Db	1196	YLTQGGGAQOPHPFPSPAFPNLTVYMODPPEERGAPSTFKGTPTAENPEYLGLOVPV	1251

```

PS      Example 3: Fig 4, 92pp; English.
XX      The present invention provides a method of treating cancer by
CC      administering a conjugate of anti-EB2 antibody with a maytensinoid. In
CC      particular, the antibody is directed against ErbB2 (also known as HER2
CC      and p185nu). The method is particularly useful in the treatment of
CC      breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC      colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ      Sequence      1255 AA;
Query Match      97.0%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;

QY      1 MELALCWMGILLALIPPGASTQVCTGTOMKRLPASPETHDMLRHLYGGCCVQGNLT 60
DB      1 MELALCWMGILLALIPPGASTQVCTGTOMKRLPASPETHDMLRHLYGGCCVQGNLT 60
QY      61 ELTYLPTNASSFLQDIOEVQGYVLAHQVQVPLRLRIVRGSTQFEDNYALAVLDNG 120
DB      61 ELTYLPTNASSFLQDIOEVQGYVLAHQVQVPLRLRIVRGSTQFEDNYALAVLDNG 120
QY      121 PLNNTTPTVTGASPGGLRELOLSLTLEILKGVTLIQNNPOLCYQDTIMKQIFHNQOLA 180
DB      121 PLNNTTPTVTGASPGGLRELOLSLTLEILKGVTLIQNNPOLCYQDTIMKQIFHNQOLA 180
QY      181 LTLIDTNSRACHPESPMCKSRCKGSESDCCSLTRTVCGAGCARCKGPLPTDCHCQC 240
DB      181 LTLIDTNSRACHPESPMCKSRCKGSESDCCSLTRTVCGAGCARCKGPLPTDCHCQC 240
QY      241 AAGCTGPRHSDCLACLPHNHSGLICELHCPALVTYNTDIPESMPNPEGRTYGASCVTACP 300
DB      241 AAGCTGPRHSDCLACLPHNHSGLICELHCPALVTYNTDIPESMPNPEGRTYGASCVTACP 300
QY      301 YNYLSTDVGSCTLVCPRLNNOEVTADGQIQRECKSKPAPRCYGLGCMQYIANKNSFGIGIT 360
DB      301 YNYLSTDVGSCTLVCPRLNNOEVTADGQIQRECKSKPAPRCYGLGCMHLEBRAVTSAN 360
QY      361 ELEPAGCKKIFGSLAFLEPESFDGDPASNTALQLOPELOVFEETLEITGYLISAMPDLSLP 420
DB      361 IOEPAGCKKIFGSLAFLEPESFDGDPASNTALQLOPELOVFEETLEITGYLISAMPDLSLP 420
QY      421 DUSVFQNTQVIRGRILNNGAYSILTQGLIGISWLGRLSIRELIGSLALIHNTHLCSFHTV 480
DB      421 DUSVFQNTQVIRGRILNNGAYSILTQGLIGISWLGRLSIRELIGSGALTIHNHTHLCFHTV 480
QY      481 PMDOLFRPNPHOALLTANRPEDECVEGELACHOLCARHCWGPGFTQCVNCSQFLRGQEC 540
DB      481 PMDOLFRPNPHOALLTANRPEDECVEGELACHOLCARHCWGPGFTQCVNCSQFLRGQEC 540
QY      541 VEECEVLQGLPREYVNAHCLPCHDECQPNQGSVTCFSPREADQCAAHYKDPFCVARC 600
DB      541 VEECEVLQGLPREYVNAHCLPCHDECQPNQGSVTCFSPREADQCAAHYKDPFCVARC 600
QY      601 BEGVKPRDLSYPMIKFPPDEBGAQCPINCHSCYDLDLDDKCPAQRASPLTSTYSAVG 660
DB      601 BEGVKPRDLSYPMIKFPPDEBGAQCPINCHSCYDLDLDDKCPAQRASPLTSTYSAVG 660
QY      661 ILVVVLGVVFGIILKRRQOKIRKTYMRLLQETELVEPLTPSGAMPQAQMRILKETEL 720
DB      661 ILVVVLGVVFGIILKRRQOKIRKTYMRLLQETELVEPLTPSGAMPQAQMRILKETEL 720
QY      721 RYKVKLGSGAFGNFTVSPWLRP-----KVASHLVEYLABNTSPKANKEILDEAYMA 775
DB      721 RYKVKLGSGAFG-----IYKGIWIPDGENVXI PVA- IKVLEBNTSPKANKEILDEAYMA 775
QY      776 GVGSPVYSRLILGICTSTYQVLTQMPGCLLDHYRENRGRGSDLLNMCQJAKGMY 835
DB      776 GVGSPVYSRLILGICTSTYQVLTQMPGCLLDHYRENRGRGSDLLNMCQJAKGMY 835
QY      836 LEDVALVHRDLAARNVLYKSPNHVKITDQGLARLLDIDETEHADGKVPFKMMALLESIL 895
DB      836 LEDVALVHRDLAARNVLYKSPNHVKITDQGLARLLDIDETEHADGKVPFKMMALLESIL 895

```

Db 836 LEDVLRHRLAARNVLKSPNHVKTIDFGLARLLDIDETEHYHAGDGKVPKIMWALESL 895
 QY RRRFTHOSDVMSTGYTWELMTFGAKPYDGI PAREIPDLLEKGRLEPPICTIDYVMIM 955
 Db 896 RRRFTHOSDVMSTGYTWELMTFGAKPYDGI PAREIPDLLEKGRLEPPICTIDYVMIM 955
 QY 956 VKCMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
 Db 956 VKCMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
 QY 1016 DIVDAEYLVPQGGFCPPDPAPGAGGMVHHRSSSTRSGGDLTLLESEBEPAPPSPL 1075
 Db 1016 DIVDAEYLVPQGGFCPPDPAPGAGGMVHHRSSSTRSGGDLTLLESEBEPAPPSPL 1075
 QY 1076 APSEGASDVEPDGDLGMAAGKLSLPTHDSPLOQRYSEDPVLEPSETDGYVAPLTCSP 1135
 Db 1076 APSEGASDVEPDGDLGMAAGKLSLPTHDSPLOQRYSEDPVLEPSETDGYVAPLTCSP 1135
 QY 1136 QPEYVNOQVPRQPSPREGELPAPAPGATLERAKTILSPKNGVWVDVAFGAVENPE 1195
 Db 1136 QPEYVNOQVPRQPSPREGELPAPAPGATLERAKTILSPKNGVWVDVAFGAVENPE 1195
 QY 1196 YLTPOGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVPV 1255
 Db 1196 YLTPOGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX 23-APR-2002 (first entry)
 DT 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX Human; ErbB; epidermal growth factor receptor; receptor;
 KM anti-ErbB antibody-maytensinoid conjugate; cancer; tumour; breast; ovary;
 KM stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KM thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KM glial disorder; astrocytal disorder; hypothalamic disorder;
 KM glandular disorder; macropagial disorder; epithelial disorder;
 KM stromal disorder; blastocoelec disorder; inflammatory disorder;
 KM angioenic disorder; immunological disorder.
 XX Homo sapiens.
 OS Homo sapiens.
 XX US2002001587-A1.
 PN 03-JAN-2002.
 PD 16-MAR-2001; 2001US-0811123.
 PF 16-MAR-2001; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX (ERIC/) ERICKSON S.
 PA (SCHM/) SCHMALL R.
 RA (SLIW/) SLIWOMSKI M.
 PI Erickson S, Schwall R, Slivkowski M,
 XX WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytensinoid conjugate to the mammal
 XX Example 3; Fig 7; 93pp; English.
 PS The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytensinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macropagial,
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.0%; Score 6608; DB 23; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;

QY 1 MELAALCRWGLLLALPPGAASVCTGTDMKRLPASPEHLDMLSHLYOGCQVVGNTL 60
 Db 1 MELAALCRWGLLLALPPGAASVCTGTDMKRLPASPEHLDMLSHLYOGCQVVGNTL 60
 QY 61 ELTYPTNASISFIQDIOEVGVYVLIANOVROYPLQRLTIVRGTCQFEDNYALAVDNG 120
 Db 61 ELTYPTNASISFIQDIOEVGVYVLIANOVROYPLQRLTIVRGTCQFEDNYALAVDNG 120
 QY 61 ELTYPTNASISFIQDIOEVGVYVLIANOVROYPLQRLTIVRGTCQFEDNYALAVDNG 120
 Db 61 ELTYPTNASISFIQDIOEVGVYVLIANOVROYPLQRLTIVRGTCQFEDNYALAVDNG 120
 QY 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYQDITLMDIFHKNQOLA 180
 Db 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYQDITLMDIFHKNQOLA 180
 QY 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYQDITLMDIFHKNQOLA 180
 Db 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYQDITLMDIFHKNQOLA 180
 QY 181 LTLIDTNRSPACHPCSPCKGSRGWSESSDDQSLTTCVACGACARCKPLPTDCHEQC 240
 Db 181 LTLIDTNRSPACHPCSPCKGSRGWSESSDDQSLTTCVACGACARCKPLPTDCHEQC 240
 QY 241 AAGCTGPRHSDCLACLFHNSGICELCPALVTNTDTPESMPPEERYTFGASCVTACP 300
 Db 241 AAGCTGPRHSDCLACLFHNSGICELCPALVTNTDTPESMPPEERYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
 QY 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
 QY 361 ELFPAGCKKIRGSLAFIPESFDGSPASNTAPLOEPQVPELTLEITGYLYISAWPDSLP 420
 Db 361 ELFPAGCKKIRGSLAFIPESFDGSPASNTAPLOEPQVPELTLEITGYLYISAWPDSLP 420
 QY 421 DLSVPQNLQVIRGRILNHGAYSLTLQGLISWLGRLRLREISGLALIHNTHTLCEVHTV 480
 Db 421 DLSVPQNLQVIRGRILNHGAYSLTLQGLISWLGRLRLREISGLALIHNTHTLCEVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHQICAGHCWGCPPTCCVNCOSQTLRQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHQICAGHCWGCPPTCCVNCOSQTLRQEC 540
 QY 541 VEECRVLOGLPREVYNARHCPCPECOQONSGVTCGPREADQVCAAHKXDPDFCVARC 600
 Db 541 VEECRVLOGLPREVYNARHCPCPECOQONSGVTCGPREADQVCAAHKXDPDFCVARC 600
 QY 601 PSQVPRDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKCPAQORASPLTISVAVG 660
 Db 601 PSQVPRDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKCPAQORASPLTISVAVG 660
 QY 661 ILVVVAVGVGILIKRQCKIRKYTMRRLLOETELVEPLTPSGAMNOMORLTKETEL 720
 Db 661 ILVVVAVGVGILIKRQCKIRKYTMRRLLOETELVEPLTPSGAMNOMORLTKETEL 720
 QY 721 RKVKVLGSGAGFNNFTVSWLRVP-----KVASHSHEVLRENTSPPRANKEILDEAVYMA 775
 Db 721 RKVKVLGSGAGFNNFTVSWLRVP-----KVASHSHEVLRENTSPPRANKEILDEAVYMA 775
 QY 776 GVGSFVYRRLGICLTISIVQLVTLQMLPRYGLLDHYHRENRRLSGDLNMCQOIAKMSY 835

```

Db      776 GVGSFVSVRLGLCLTSTVQLTQMLPGCLLDHVRNRGLSGDILLNMCQIAKMSY 835
Qy      836 LEDVRLVHRDLAARVLYKSPNHVITPFGARLLIDETEHADGGKVPKXMALESIL 895
Db      836 LEDVRLVHRDLAARVLYKSPNHVITPFGARLLIDETEHADGGKVPKXMALESIL 895
Qy      896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKEGERLPQPPCTIDVYIM 955
Db      896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKEGERLPQPPCTIDVYIM 955
Qy      956 VKCWMIDSECRPFRELVESEFSRMAPDQRFVYVIONEDLGPASPLDSFTFYSLEDDMG 1015
Db      956 VKCWMIDSECRPFRELVESEFSRMAPDQRFVYVIONEDLGPASPLDSFTFYSLEDDMG 1015
Qy      1016 DLVDAEEYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDDLTGLSEPEEAPRSPL 1075
Db      1016 DLVDAEEYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDDLTGLSEPEEAPRSPL 1075
Qy      1076 ABEGAGSDVFPDGDGMAKGLSLPTHDPSPLORYSEDPVLPSETDGYVAPLTCSP 1135
Db      1076 ABEGAGSDVFPDGDGMAKGLSLPTHDPSPLORYSEDPVLPSETDGYVAPLTCSP 1135
Qy      1136 QPEYVNPDPVRPQPSPREGRLPAARPAATLERAKTLPQNGVXKDVAFGAVENPE 1195
Db      1136 QPEYVNPDPVRPQPSPREGRLPAARPAATLERAKTLPQNGVXKDVAFGAVENPE 1195
Qy      1196 YLTPQGAAPQHPHPAPSPAFDNLTYWDQDPPEGAPSPSTFKGPTANENPEYGLDVPV 1255
Db      1196 YLTPQGAAPQHPHPAPSPAFDNLTYWDQDPPEGAPSPSTFKGPTANENPEYGLDVPV 1255

RESULT 5
AA001111
ID AA001111 standard; protein; 1255 AA.
XX
AC AA001111;
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KM HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KM breast cancer; ovary cancer; colon cancer; lung cancer;
KM prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
FN WO9630514-A1.
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW ) UNIV WASHINGTON.
PI Cheever MA, Disis ML;
XX
DR WPI; 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.

```

```

XX
CC Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;

```

```

Query Match 96.9%; Score 6602; DB 17; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

```

```

Qy      1 MELALCEWGLLALLPPGAASVQCTGTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Db      1 MELALCEWGLLALLPPGAASVQCTGTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Qy      61 ELTYIPTASLSFLDIDQENQGVYLIANOVROYPLRLIRVGTQLPEDNYALAVLDNG 120
Db      61 ELTYIPTASLSFLDIDQENQGVYLIANOVROYPLRLIRVGTQLPEDNYALAVLDNG 120
Qy      121 DPLNNTPTVPGASPGGLREIQRLSLTEILKGVLIQENPOLCYQDTILMKDIFHKNNQLA 180
Db      121 DPLNNTPTVPGASPGGLREIQRLSLTEILKGVLIQENPOLCYQDTILMKDIFHKNNQLA 180
Qy      181 LTLIDTNRBRACHPCSPCKGSRCKWSESSDCCSLTRVACGACARCGPLPTCCHEQC 240
Db      181 LTLIDTNRBRACHPCSPCKGSRCKWSESSDCCSLTRVACGACARCGPLPTCCHEQC 240
Qy      241 AAGCTGPKXSDCLACLNHNSGICELHCPALVTYNTDFEEMPNPEGRYTFGASCVTACP 300
Db      241 AAGCTGPKXSDCLACLNHNSGICELHCPALVTYNTDFEEMPNPEGRYTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGNOYIKANSKFTGIT 360
Db      301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGNOYIKANSKFTGIT 360
Qy      361 ELFPAGCKKIFGSLAFLESPFPDQASNTATPLOPOLQVETLEETGYLYISAMPDLP 420
Db      361 IDEFACCKKIFGSLAFLESPFPDQASNTATPLOPOLQVETLEETGYLYISAMPDLP 420
Qy      421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNTTHLCFVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNTTHLCFVHTV 480
Qy      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQLCARHGMGPPTQCVNCSOFLRQEC 540
Db      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQLCARHGMGPPTQCVNCSOFLRQEC 540
Qy      541 VEECRVLOGLPREYVNAARHCLPEHECOFQNGSVTCFPEADQCVACAHYDPPFCVARC 600
Db      541 VEECRVLOGLPREYVNAARHCLPEHECOFQNGSVTCFPEADQCVACAHYDPPFCVARC 600
Qy      601 PSGVRPDLISYMPIWKFPDEBGAQCPCPINCTHSQVLDKDCGAPQASPLTSYSAVYG 660
Db      601 PSGVRPDLISYMPIWKFPDEBGAQCPCPINCTHSQVLDKDCGAPQASPLTSYSAVYG 660
Qy      661 ILLVVLGVVPGILIKRQOKIRKXTMRLLQETLVEPLTPSGAMPQAOIRLIKETEL 720
Db      661 ILLVVLGVVPGILIKRQOKIRKXTMRLLQETLVEPLTPSGAMPQAOIRLIKETEL 720
Qy      721 RKYKVLGSGAFGFPNNFTVSFWLRV-----KYASLHEVLRENTSPKANKELDEAYYMA 775
Db      721 RKYKVLGSGAFG---TYKGIWIPDEGEVAKIPVA-IKYLRENTSPKANKELDEAYYMA 775
Qy      776 GVGSFVSVRLGLCLTSTVQLTQMLPGCLLDHVRNRGLSGDILLNMCQIAKMSY 835
Db      776 GVGSFVSVRLGLCLTSTVQLTQMLPGCLLDHVRNRGLSGDILLNMCQIAKMSY 835

```


QY 836 LEDVRLVHRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGKVPDKMMALESIL 895
 DB 836 LEDVRLVHRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGKVPDKMMALESIL 895
 QY 896 RRRFTHQSDVMSYGVTTVMELMTFGAKPYDGI:PAEIPDLLEKSGRLNQPPICITIDVYIM 955
 DB 896 RRRFTHQSDVMSYGVTTVMELMTFGAKPYDGI:PAEIPDLLEKSGRLNQPPICITIDVYIM 955
 QY 956 VECWVIDSECRRPFEIYSEFSRMAKDQRFVVIQNEDLGASPLDSTFYRSLLDDDMG 1015
 DB 956 VECWVIDSECRRPFEIYSEFSRMAKDQRFVVIQNEDLGASPLDSTFYRSLLDDDMG 1015
 QY 1016 DLVDAEEYLVPOQGFPCDPAAGAGWVHRHRSSTSTRSGGDI:TLGLEPSEEEAPRSPL 1075
 DB 1016 DLVDAEEYLVPOQGFPCDPAAGAGWVHRHRSSTSTRSGGDI:TLGLEPSEEEAPRSPL 1075
 QY 1076 ABSEGASDVFDGDLGMAAGKLOSLPTHDSPLQRYSEDPVLPBSETGYVAPLTCSP 1135
 DB 1076 ABSEGASDVFDGDLGMAAGKLOSLPTHDSPLQRYSEDPVLPBSETGYVAPLTCSP 1135
 QY 1136 QPEYVNPDPVRPQPSPREGSLPAARPAATLERAKTLPQKNGVVDVPAFGGAVENPE 1195
 DB 1136 QPEYVNPDPVRPQPSPREGSLPAARPAATLERAKTLPQKNGVVDVPAFGGAVENPE 1195
 QY 1196 YLTPQGAAPQHPAPPASPAFDNIYYNDQDPPERGAPESTFKGTPTAENPEYIGLDVPV 1255
 DB 1196 YLTPQGAAPQHPAPPASPAFDNIYYNDQDPPERGAPESTFKGTPTAENPEYIGLDVPV 1255

RESULT 6
 AAM92406
 ID AAM92406 standard; Protein; 1255 AA.
 AC AAM92406;
 XX
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 XX
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 XX malignancy; treatment; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 676..1255
 FT /note= "region which elicits immune response"
 XX
 PN US869445-A.
 XX
 PD 09-FEB-1999.
 XX
 PF 01-APR-1996; 96US-0625101.
 XX
 PR 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-003644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 PA (UNIT) UNIT WASHINGTON.
 XX
 PI Cheever MA, Disis ML;
 XX
 DR MPI: 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 PS Claim 3; Column 31-38; 26pp; English.
 XX
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or recurrence.
 CC
 XX

SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 20; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCRMGILLALPPGAASVCTGTDMKRLPASPEHIDMLRHLYQCCVQGNL 60
 DB 1 MELALCRMGILLALPPGAASVCTGTDMKRLPASPEHIDMLRHLYQCCVQGNL 60
 QY 61 ELTYLPTNASLSPLODIOEVGYVLIANOVROVPLQRLRVSTQLEEDVYALAVDNG 120
 DB 61 ELTYLPTNASLSPLODIOEVGYVLIANOVROVPLQRLRVSTQLEEDVYALAVDNG 120
 QY 121 DELNNTTPTGASPGGLREQLRLTEILKGVLIQRPOLCYODTILMKDIFKHNNOLA 180
 DB 121 DELNNTTPTGASPGGLREQLRLTEILKGVLIQRPOLCYODTILMKDIFKHNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPKCKSRCKWSESSDDCSLTRTVCAGGCARCKPLPTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPKCKSRCKWSESSDDCSLTRTVCAGGCARCKPLPTDCHEQC 240
 QY 241 AAGCTGPHNSDCLAFHNHSGICELHCPALVTYNTDTFESMPREERYTFGASCTVAC 300
 DB 241 AAGCTGPHNSDCLAFHNHSGICELHCPALVTYNTDTFESMPREERYTFGASCTVAC 300
 QY 301 YNYLSTDVSGCTLVCPHINQEVTAEDGTQRCCKSKPCARVYGLGNOYIKANSKFTGIT 360
 DB 301 YNYLSTDVSGCTLVCPHINQEVTAEDGTQRCCKSKPCARVYGLGNOYIKANSKFTGIT 360
 QY 361 ELTFAGCKKTFGSLAFLEPSFDGPASTAPLOEQOVFTLEITGYLYISMPDLSL 420
 DB 361 ELTFAGCKKTFGSLAFLEPSFDGPASTAPLOEQOVFTLEITGYLYISMPDLSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLIGISMLGRSLRELSGLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLIGISMLGRSLRELSGLALIHNNTHLCFVHTV 480
 QY 481 PWDQFLRNPQALHTTANRPEDECVGBGLACHQICARHCKGPPPTQCVNCSOFLRQEC 540
 DB 481 PWDQFLRNPQALHTTANRPEDECVGBGLACHQICARHCKGPPPTQCVNCSOFLRQEC 540
 QY 541 VECRVLOGLPREYVNAHCLPCHPECCPONGSVTCGPEADQCAAHYDPPFCVARC 600
 DB 541 VECRVLOGLPREYVNAHCLPCHPECCPONGSVTCGPEADQCAAHYDPPFCVARC 600
 QY 601 PSGVXPDLISYPIWKFPDEBACQPCPINCTHSCVDLDKGCPEAQBASPSTISVAVVG 660
 DB 601 PSGVXPDLISYPIWKFPDEBACQPCPINCTHSCVDLDKGCPEAQBASPSTISVAVVG 660
 QY 661 ILLVVVLGVNREGILIKRQOKIRKXTVRRLOEELVBPILPSGAMNOQMRILKXTEL 720
 DB 661 ILLVVVLGVNREGILIKRQOKIRKXTVRRLOEELVBPILPSGAMNOQMRILKXTEL 720
 QY 721 RKVKYLGSGAGFNNFTVSFWLRVP----KVSASHLEVIRENTSPRANKEILDEAYVMA 775
 DB 721 RKVKYLGSGAGF---TYKGIWIPDGENVXIPVA-IKVLRENTSPRANKEILDEAYVMA 775
 QY 776 GVGSPYVSRILGICLITSVQLVTQLMFPGCLLDHYRNRKGLSGQDILNMCQIAKMSY 835
 DB 776 GVGSPYVSRILGICLITSVQLVTQLMFPGCLLDHYRNRKGLSGQDILNMCQIAKMSY 835
 QY 836 LEDVRLVHRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGKVPDKMMALESIL 895
 DB 836 LEDVRLVHRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGKVPDKMMALESIL 895

```
QY 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
XX |||||
DB 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
QY 956 VKCWMIDSECRPRFRELVSFESRMAADPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
DB 956 VKCWMIDSECRPRFRELVSFESRMAADPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
QY 1016 DLVDABEYLVPOQGFPCPPAPAGAGMHHRRSSSTSGGDLTLTGLEPSEEPAPSP 1075
DB 1016 DLVDABEYLVPOQGFPCPPAPAGAGMHHRRSSSTSGGDLTLGLEPSEEPAPSP 1075
QY 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVALTCSP 1135
DB 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVALTCSP 1135
QY 1136 QPEYVNOQPDVPRPSPSPRSGPLPAAPAPATLERKATLSFGKNGVYKQVFAFGAVENPE 1195
DB 1136 QPEYVNOQPDVPRPSPSPRSGPLPAAPAPATLERKATLSFGKNGVYKQVFAFGAVENPE 1195
QY 1196 YLTPQGAAPQPHPPAPAFSPAFNDLYWQDPPERGAAPPSTFKGTPTAENPEYLGLDV 1255
DB 1196 YLTPQGAAPQPHPPAPAFSPAFNDLYWQDPPERGAAPPSTFKGTPTAENPEYLGLDV 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KM Human; HER-2/neu; oncogene; tyrosine kinase; cytostratic; vaccine;
KM breast cancer; prostate cancer; ovarian cancer; lung cancer;
KM colon cancer.
XX
OS Homo sapiens.
XX
PN M0200044899-A1.
PD 03-AUG-2000.
XX
PE 28-JAN-2000; 2000MO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-P3DB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
```

```
CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 21; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MBLAALCRMGILLALLPFGAASCTGTGDKLRLPASPETHLMLRLHYOGCGVVGNTL 60
DB 1 MBLAALCRMGILLALLPFGAASCTGTGDKLRLPASPETHLMLRLHYOGCGVVGNTL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVIDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVIDNG 120
QY 121 DPLNNTPTPTGASPGALRELRSLTEILKGGVLIQRNPOLCYQDTILMKQIFPKNNOLA 180
DB 121 DPLNNTPTPTGASPGALRELRSLTEILKGGVLIQRNPOLCYQDTILMKQIFPKNNOLA 180
QY 181 LTLIDTNSRACHPCSPMCKGSRGCESSSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNSRACHPCSPMCKGSRGCESSSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
QY 241 AAGCTGPKASDCLACLRHNSGICELHCPALVYNTDTFESMPNPEGRTTGAACVTACP 300
DB 241 AAGCTGPKASDCLACLRHNSGICELHCPALVYNTDTFESMPNPEGRTTGAACVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHNOEYTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPLNHNOEYTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
QY 361 ELEFAGCKKXIFGSLAFLEPSEFDGDPASNTAPLQEPQLQVFETLEITGYLYISAMPDLP 420
DB 361 IGFACCKKXIFGSLAFLEPSEFDGDPASNTAPLQEPQLQVFETLEITGYLYISAMPDLP 420
QY 421 DLSVQNLQVIRGRLIHNGAVSLTLOGIGISWLRISREJSGSLALIHNTHLCFYHTV 480
DB 421 DLSVQNLQVIRGRLIHNGAVSLTLOGIGISWLRISREJSGSLALIHNTHLCFYHTV 480
QY 481 PMDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARHCWGPPTQCVCNSQFLRGEC 540
DB 481 PMDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARHCWGPPTQCVCNSQFLRGEC 540
QY 541 VEECRVLQGLPREYNNARCLFCHECQOPQNSYTCFEPBADQCACAHYKDPFCVARC 600
DB 541 VEECRVLQGLPREYNNARCLFCHECQOPQNSYTCFEPBADQCACAHYKDPFCVARC 600
QY 601 PSGVKRDLSTYMPIMKFPDEGACQCPINCTHSQVLDLDCGCPAQRASPLTISIVAVG 660
DB 601 PSGVKRDLSTYMPIMKFPDEGACQCPINCTHSQVLDLDCGCPAQRASPLTISIVAVG 660
QY 661 ILIVVVLGVFGILIKRQOKIRKTYMRRLQETLVEPLTPSGAMPQAOQMRILKETEL 720
DB 661 ILIVVVLGVFGILIKRQOKIRKTYMRRLQETLVEPLTPSGAMPQAOQMRILKETEL 720
QY 721 RKVKVLGSAFGFNNTVSFWLRVP----KVSASHLEVRNTPSKXNKELIDEAAYMA 775
DB 721 RKVKVLGSAFGFNNTVSFWLRVP----KVSASHLEVRNTPSKXNKELIDEAAYMA 775
QY 776 GVGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVRNRRGRLGSDLLNMCQIAKGSY 835
DB 776 GVGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVRNRRGRLGSDLLNMCQIAKGSY 835
QY 836 LEBVRVLVHMDLAARNLVKSPNHVITPBGRLALDIDETEYHADGQVPIKMALESIL 895
DB 836 LEBVRVLVHMDLAARNLVKSPNHVITPBGRLALDIDETEYHADGQVPIKMALESIL 895
QY 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
DB 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
QY 956 VKCWMIDSECRPRFRELVSFESRMAADPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
```

DB 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEIDLGPASFLDSTFYRSLLDDMG 1015
QY 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075
DB 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075
QY 1076 APSEGAGSDVDFDGLGMAAKGLSLPTHDSPLOQYSEDPVPLPSETDGYVAPLTCSP 1135
DB 1076 APSEGAGSDVDFDGLGMAAKGLSLPTHDSPLOQYSEDPVPLPSETDGYVAPLTCSP 1135
QY 1136 QPEYVNOQDVPRPPSPREGPLPAPAPAGATLEAKTLSPGKGVVVDVAFGAVENPE 1195
DB 1136 QPEYVNOQDVPRPPSPREGPLPAPAPAGATLEAKTLSPGKGVVVDVAFGAVENPE 1195
QY 1196 YLTPGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLGLDVPY 1255
DB 1196 YLTPGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLGLDVPY 1255

RESULT 8
AAy84780
ID AAy84780 standard; Protein; 1255 AA.
XX
AC AAy84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPICE erbB-2 receptor protein.
XX
KM SPICE erbB-2 receptor protein; cell transformation disorder; cancer;
KM tumor cell proliferation; tissue degeneration; arthropathy;
KM bone resorption; inflammatory disease; degenerative disorder;
KM wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMMASTER.
PI Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as atrophy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 21; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
QY 1 MELALCMWGLLLALPPGASTVCTGTMKRLPAPSPETHLDMRLHYOGGVQVGNL 60
DB 1 MELALCMWGLLLALPPGASTVCTGTMKRLPAPSPETHLDMRLHYOGGVQVGNL 60
QY 61 ELYLPTNASLSPLODIOEVQVYLIANOVQVFLQRLRIVRGTOLFEDNYALAVLNG 120
DB 61 ELYLPTNASLSPLODIOEVQVYLIANOVQVFLQRLRIVRGTOLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVTAASGGLRELOBSLTLLKGVYLIQNPOLCTQDTLLMDIFHKNNQLA 180
DB 121 DPLNNTTPTVTAASGGLRELOBSLTLLKGVYLIQNPOLCTQDTLLMDIFHKNNQLA 180
QY 181 LTLIDTNSRACHSPKSPKCSRWCSESSDCCSLTRTYVAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDTNSRACHSPKSPKCSRWCSESSDCCSLTRTYVAGGACARCKPLPTDCHEQC 240
QY 241 AAGCTGPHGSDCLALCPHNSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCYTAAP 300
DB 241 AAGCTGPHGSDCLALCPHNSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCYTAAP 300
QY 301 YNYLSTDVSGCTLVCPHNOEVTAEQTCRCKSPCAVCTGLGHEHLREVAATYSAN 360
DB 301 YNYLSTDVSGCTLVCPHNOEVTAEQTCRCKSPCAVCTGLGHEHLREVAATYSAN 360
QY 361 ELEFAGCKKIGSLAFIPESFDGPASNTAPLOPEQLOVEETEIEITGYLISAMPDLP 420
DB 361 IQEFAGCKKIGSLAFIPESFDGPASNTAPLOPEQLOVEETEIEITGYLISAMPDLP 420
QY 421 DLSVFQNIQVIRGRILNHGAYSLTLOGLIGSLGRLBELGSLALIHNTHLCPHTV 480
DB 421 DLSVFQNIQVIRGRILNHGAYSLTLOGLIGSLGRLBELGSLALIHNTHLCPHTV 480
QY 481 PMDOLFRPHQALHTANRPEDECVGEGALCHQACARHGMCPPTCCVNCOSFLRQEC 540
DB 481 PMDOLFRPHQALHTANRPEDECVGEGALCHQACARHGMCPPTCCVNCOSFLRQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600
DB 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600
QY 601 PSQVPRDLSVMPIMKFPDEBGACQPCINCHSCVDLDDGCPABQASPLTISAVVG 660
DB 601 PSQVPRDLSVMPIMKFPDEBGACQPCINCHSCVDLDDGCPABQASPLTISAVVG 660
QY 661 ILVVVLGVVFGILIKRQCKIRKYTRRLQETELVEPLTPSGAMPNQAOMRLKETEL 720
DB 661 ILVVVLGVVFGILIKRQCKIRKYTRRLQETELVEPLTPSGAMPNQAOMRLKETEL 720
QY 721 RKVKYLSGAGFNNFVFWLRVP-----KVASHLSEVLRRENTSPKANEILDEAVMA 775
DB 721 RKVKYLSGAGFNNFVFWLRVP-----KVASHLSEVLRRENTSPKANEILDEAVMA 775
QY 776 GVGSPYVSRLLGICLTSTVQVLTOLMPYGLLDHVRNRRGLSGODLLNMCQIAKMSY 835
DB 776 GVGSPYVSRLLGICLTSTVQVLTOLMPYGLLDHVRNRRGLSGODLLNMCQIAKMSY 835
QY 836 LEDVRLVARDIAANVILKSPNHVKTDFGLARLLIDETBYHADGKVPDKMALESIL 895
DB 836 LEDVRLVARDIAANVILKSPNHVKTDFGLARLLIDETBYHADGKVPDKMALESIL 895
QY 896 RRRFTHQSDWSYGTWELMTFAKRYDGPAPRIEDLLEKGRJPOPICTIDVYIM 955
DB 896 RRRFTHQSDWSYGTWELMTFAKRYDGPAPRIEDLLEKGRJPOPICTIDVYIM 955
QY 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEIDLGPASFLDSTFYRSLLDDMG 1015
DB 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEIDLGPASFLDSTFYRSLLDDMG 1015
QY 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075

DB 1016 DLVDAEEYLVPQGGFFCCDDPAPAGAGMWHHRSSSTRSGGDDTLTGLFSEBEAPRSP 1075
QY 1076 APSEAGSDVDFDGLGMAAGKGLSLPHDPSPLQRYSEDDPTVLPSEDTGYVAPLTCS 1135
DB 1076 APSEAGSDVDFDGLGMAAGKGLSLPHDPSPLQRYSEDDPTVLPSEDTGYVAPLTCS 1135
QY 1136 QPEYVNOPDVRPOPSPREGPLPARPAGATLERAKTSLPGKNGVVDVAFGAVENPE 1195
DB 1136 QPEYVNOPDVRPOPSPREGPLPARPAGATLERAKTSLPGKNGVVDVAFGAVENPE 1195
QY 1196 YLTPQGAAPQHPAPSPAPFADNLVWDQDPPERGAPSPSTFKGTPAENPEYLGLDVPV 1255
DB 1196 YLTPQGAAPQHPAPSPAPFADNLVWDQDPPERGAPSPSTFKGTPAENPEYLGLDVPV 1255
RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytosolic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX
PI Cheever MA, Hand-Zimmermann S;
XX
DR MPI: 2001-476112/51.
XX
DR N-PSDB; AAH23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 22; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

DB 61 ELTYLPTNASTSLFQDIQEVGYVLIANQVQVPLQRLRIYRGTLFEDNYALAVLDNG 120
QY 121 DPLANTTYTASRGGLRELOLRLETLIKGVYLIGNPOLCTQDTILMMDIHHKNOQA 180
DB 121 DPLANTTYTASRGGLRELOLRLETLIKGVYLIGNPOLCTQDTILMMDIHHKNOQA 180
QY 181 LTLIDTNSRACPCSPCKSPCKSPCKSGSESDCCSLTRTVACGACGACRCKPLPTDCHEQC 240
DB 181 LTLIDTNSRACPCSPCKSPCKSPCKSGSESDCCSLTRTVACGACGACRCKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDTFESMPNBSGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDTFESMPNBSGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTVQPLNCEVTAEQDQRECKSKRCARVCGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVSGCTVQPLNCEVTAEQDQRECKSKRCARVCGLGMQYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPELOQVEFTELEITGYLYISAMPDLP 420
DB 361 ELEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPELOQVEFTELEITGYLYISAMPDLP 420
QY 421 DLSVFQNTQVIRGRILHNGAYSLLQGLISWGLRLSRLRELSGLALIHNTHLCFVHTV 480
DB 421 DLSVFQNTQVIRGRILHNGAYSLLQGLISWGLRLSRLRELSGLALIHNTHLCFVHTV 480
QY 481 PMDQLEFRPHQALHTANRPEDECVGEGLACHQLCARHGCMGPPTQCVNCSQFLRQEC 540
DB 481 PMDQLEFRPHQALHTANRPEDECVGEGLACHQLCARHGCMGPPTQCVNCSQFLRQEC 540
QY 541 VEECVVLQGLPREYVNAHCLPCHPECOQPNQSVTCTGPEADQCVACAHKXDPFCVARC 600
DB 541 VEECVVLQGLPREYVNAHCLPCHPECOQPNQSVTCTGPEADQCVACAHKXDPFCVARC 600
QY 601 PSGVAPDLSTYMPIMKFPDEBACQPCPINCCHSCVDDDDKQCPAEQASPLTSLVSAVWG 660
DB 601 PSGVAPDLSTYMPIMKFPDEBACQPCPINCCHSCVDDDDKQCPAEQASPLTSLVSAVWG 660
QY 661 ILVVVLGVGFILIKRQOKIRKYMRLLETELVEPLTPSGAMPNOQMRILKETEL 720
DB 661 ILVVVLGVGFILIKRQOKIRKYMRLLETELVEPLTPSGAMPNOQMRILKETEL 720
QY 721 RKVXLTGSGAGFNNFTYSFWLRP-----VVSASHLEVRENTSPRANELLIDEAVYMA 775
DB 721 RKVXLTGSGAGFNNFTYSFWLRP-----VVSASHLEVRENTSPRANELLIDEAVYMA 775
QY 776 GVGSFYVSRLTGLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLLNMCQIAKQMSY 835
DB 776 GVGSFYVSRLTGLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLLNMCQIAKQMSY 835
QY 836 LEDVRLVNRDLAARVVLKSPNHYKITDFGLARLLDDETYHADGSKVPIKMALESIL 895
DB 836 LEDVRLVNRDLAARVVLKSPNHYKITDFGLARLLDDETYHADGSKVPIKMALESIL 895
QY 896 RRRFTHQSDWMSYGTWELMTFGAKPYDGIIPAREIDPLEKGRLELPQPICTIDVYIM 955
DB 896 RRRFTHQSDWMSYGTWELMTFGAKPYDGIIPAREIDPLEKGRLELPQPICTIDVYIM 955
QY 956 VKCMMIDSECRPFELVSEFSRMAARDQRRVVIQNDLGPASPLDSTFRSLLEDDMG 1015
DB 956 VKCMMIDSECRPFELVSEFSRMAARDQRRVVIQNDLGPASPLDSTFRSLLEDDMG 1015
QY 1016 DLVDAEEYLVPQGGFFCCDDPAPAGAGMWHHRSSSTRSGGDDTLTGLFSEBEAPRSP 1075
DB 1016 DLVDAEEYLVPQGGFFCCDDPAPAGAGMWHHRSSSTRSGGDDTLTGLFSEBEAPRSP 1075
QY 1076 APSEAGSDVDFDGLGMAAGKGLSLPHDPSPLQRYSEDDPTVLPSEDTGYVAPLTCS 1135
DB 1076 APSEAGSDVDFDGLGMAAGKGLSLPHDPSPLQRYSEDDPTVLPSEDTGYVAPLTCS 1135
QY 1136 QPEYVNOPDVRPOPSPREGPLPARPAGATLERAKTSLPGKNGVVDVAFGAVENPE 1195

DB 1136 QPEYVNCPEVRPQPPSPREGCLPAPAPAGATLERPKTLSPGKNGVWDVFAFGAVENPE 1195
 QY 1196 YLTPOGGAPOHPHPAPSPAFDNLXYMDODPPERGAAPSTFKGTPTAENPEYLGLDVPV 1255
 DB 1196 YLTPOGGAPOHPHPAPSPAFDNLXYMDODPPERGAAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.

AC AAG88267;
 DT 11-SEP-2001 (first entry)
 DE HER2/neu amino acid sequence.
 KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KW Immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
 KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 OS Homo sapiens.
 PN MO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000MO-US33591.
 PR 10-DEC-1999; 99US-0458299.
 PI (EPTM-) EPIMONE INC.
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 PI Keogh E;
 DR WPI: 2001-374995/39.
 PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 PS Disclosure; Page 15; 1999p; English.
 XX

CC The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (I), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.
 CC
 CC Sequence 1255 AA;
 SQ

Query Match 96.9%; Score 6602; DB 22; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCWGILLILPFGASTVCGTGMKRLPASPETHLDMLHLYOGCGVGVGNL 60
 DB 1 MELALCWGILLILPFGASTVCGTGMKRLPASPETHLDMLHLYOGCGVGVGNL 60

QY 61 ELTYLPTNASLFLDIOEVGYVLIANQVRYVPLQRLIRVGTQLFEDNALAVLDNG 120
 DB 61 ELTYLPTNASLFLDIOEVGYVLIANQVRYVPLQRLIRVGTQLFEDNALAVLDNG 120

QY 121 PLNNTTPTVTGASPGGLRELOLRLTELKGGVLIQNNPOLCYDTLLMDIFPKNNOLA 180
 DB 121 PLNNTTPTVTGASPGGLRELOLRLTELKGGVLIQNNPOLCYDTLLMDIFPKNNOLA 180

QY 181 LTLIDTNSRACHPSPCKSGSRGMSSEDCQSLTRTVGAGCARCKGPLPTDCHEQC 240
 DB 181 LTLIDTNSRACHPSPCKSGSRGMSSEDCQSLTRTVGAGCARCKGPLPTDCHEQC 240

QY 241 AAGCTGPRHSDCLALPHNHSIGICELCPALVTYNTDFESMPNREGRYTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLALPHNHSIGICELCPALVTYNTDFESMPNREGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGCTIVGPIHNCETVEDGTQCEKSKPCARVCYGLMGQYIKANSKFGIT 360
 DB 301 YNYLSTDVSGCTIVGPIHNCETVEDGTQCEKSKPCARVCYGLMGQYIKANSKFGIT 360

QY 361 ELEFAGCKKIFGSLAFPLESFEGDPASNTAPLOEQLQVETLEITGYLYISAMPDPL 420
 DB 361 IEFAGCKKIFGSLAFPLESFEGDPASNTAPLOEQLQVETLEITGYLYISAMPDPL 420

QY 421 DLSVQNIQVIRGRILNHGAYSLTQIGISWGLRLRLRELGSLALIHNTHLCPHNTV 480
 DB 421 DLSVQNIQVIRGRILNHGAYSLTQIGISWGLRLRLRELGSLALIHNTHLCPHNTV 480

QY 481 PWDQLFRRPHQALHTANRPEDECVGEGLACHQCARGHCGSPPTCCVNCSPFLRGQC 540
 DB 481 PWDQLFRRPHQALHTANRPEDECVGEGLACHQCARGHCGSPPTCCVNCSPFLRGQC 540

QY 541 VEECVLQGLPREYVNAHCLPHECQPNQSVTCFGEADQCVCACAHYKDPFCVARC 600
 DB 541 VEECVLQGLPREYVNAHCLPHECQPNQSVTCFGEADQCVCACAHYKDPFCVARC 600

QY 601 PSGVPRDLSWPIKPFPEBEGACQPCPNCHSGVDLDDKGCPEQASPLTSLVSVNG 660
 DB 601 PSGVPRDLSWPIKPFPEBEGACQPCPNCHSGVDLDDKGCPEQASPLTSLVSVNG 660

QY 661 ILVVVLGVVFGILIKRQKIRRYTWRLQLQETELVEPLTPSGAMENQAMRLKETEL 720
 DB 661 ILVVVLGVVFGILIKRQKIRRYTWRLQLQETELVEPLTPSGAMENQAMRLKETEL 720

QY 721 RKVKYLSGAGFNNFTVSEFWLRVP-----KVASHLVELRNTSPANKKEILDEAYMA 775
 DB 721 RKVKYLSGAGFNNFTVSEFWLRVP-----KVASHLVELRNTSPANKKEILDEAYMA 775

QY 776 GVGSPYVRLGICITSTVQVLVLTQMPYGLDLVHRENRGLSGQDLNNCMQJAKMSY 835
 DB 776 GVGSPYVRLGICITSTVQVLVLTQMPYGLDLVHRENRGLSGQDLNNCMQJAKMSY 835

QY 836 LEDVLRVHRDLAARNVLYKSPNHXYITDFGLARLLDDEHYHADGKVPDKMALESIL 895
 DB 836 LEDVLRVHRDLAARNVLYKSPNHXYITDFGLARLLDDEHYHADGKVPDKMALESIL 895

QY 896 RRRPFGHSDWSYGTWELMTFGAKPYDGIPAEIPDLLEKGRLEPPICITIDVIMM 955
 DB 896 RRRPFGHSDWSYGTWELMTFGAKPYDGIPAEIPDLLEKGRLEPPICITIDVIMM 955

QY 956 VKCWMIDSECRPRRELIVSEFSRMAPDRFVVIQNEIDLPAFLDSTFYRSLLEDDMG 1015
 DB 956 VKCWMIDSECRPRRELIVSEFSRMAPDRFVVIQNEIDLPAFLDSTFYRSLLEDDMG 1015

QY 1016 DLVDAEELVPOQGFECDDPARAGWVHRHRSSSTRSGGDTLTLGSEBEPAPRSL 1075
 DB 1016 DLVDAEELVPOQGFECDDPARAGWVHRHRSSSTRSGGDTLTLGSEBEPAPRSL 1075

```
Db      1016 DLVDAEYLVPQGGFCFPDPAAPGAGGVHHRHSSSTRSGGDLTLGLPSEEBEAPRPL 1075
Qy      1076 APSEGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPBETGGYAAPLTGSP 1135
Db      1076 APSEGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPBETGGYAAPLTGSP 1135
Qy      1136 QPEYVNOQDVPRQPPSPREGPLPAARFAGATLIERAKTILSPGKNGVVDVAFGAVENPE 1195
Db      1136 QPEYVNOQDVPRQPPSPREGPLPAARFAGATLIERAKTILSPGKNGVVDVAFGAVENPE 1195
Qy      1196 YLTPQGAAPQHPPPAPSPAFDNLVYWDQDPPEBGAPESTFKGTPTAENPEYLGLDVPV 1255
Db      1196 YLTPQGAAPQHPPPAPSPAFDNLVYWDQDPPEBGAPESTFKGTPTAENPEYLGLDVPV 1255

RESULT 11
AAE24067
ID      AAE24067 standard; Protein; 1255 AA.
XX
AC      AAE24067;
XX
DT      23-SEP-2002 (first entry)
XX
DE      Human Her-2 protein.
XX
KW      Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KM      hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW      tumour; gene therapy; phosphorothioate backbone.
XX
OS      Homo sapiens.
XX
PN      WO200222636-A1.
XX
PD      21-MAR-2002.
XX
PF      12-SEP-2001; 2001WO-US28572.
XX
PR      15-SEP-2000; 2000US-0663834.
XX
PA      (ISIS-) ISIS PHARM INC.
XX
PI      Bennett CF, Cowseart LM;
XX
DR      WPI; 2002-471192/50.
XX
DR      N-PSDB; AAD38904.
XX
PT      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT      inflammation or to prevent infection in humans -
XX
XX      Example 13; Page 95-107; 116pp; English.
XX
CC      The invention relates to antisense compounds targeted to a nucleic
CC      acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC      that specifically hybridises with and inhibits the expression of Her2.
CC      Antisense compounds of the invention are used for treating diseases or
CC      conditions associated with Her2 such as hyperproliferative disorders
CC      e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC      neural or cardiac cancer. They are also useful prophylactically e.g.
CC      to prevent or delay infection, inflammation and tumour formation. The
CC      invention is also used in gene therapy. The present sequence is human
CC      Her-2 protein.
XX
SQ      Sequence      1255 AA;

Query Match      96.9%; Score 6602; DB 23; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0.
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

Qy      1 MELAALCRWGLLALLPGGAATQVCTGDMKRLRPASPEHLDMLRHLVGGCGVQGNL 60
Db      1 MELAALCRWGLLALLPGGAATQVCTGDMKRLRPASPEHLDMLRHLVGGCGVQGNL 60

Qy      1136 QPEYVNOQDVPRQPPSPREGPLPAARFAGATLIERAKTILSPGKNGVVDVAFGAVENPE 1195
```

|||||
Db 1136 QPEYVQPDVAPPPSPRECEPLPAPAPGATLERPKLSPKGNVADVAFGAVENPE 1195
QY 1136 YLTPOGGAAPQHPPPAPSPAFDNLYVDDPPRRGAPSPSTFKCTPTAENPEYLGLDVPV 1255
Db 1136 YLTPOGGAAPQHPPPAPSPAFDNLYVDDPPRRGAPSPSTFKCTPTAENPEYLGLDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
XX
XX 01-JUL-2002 (first entry)
DT
XX
XX Human Her-2/neu protein.
DE
XX Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note="Naturally processed HLA-B44-restricted epitope"
XX
XX WO200214503-A2.
PN
XX 21-FEB-2002.
PD
XX 14-AUG-2001; 2001MO-US41733.
PF
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Hand-Zimmermann S, Cheever MA, Foy TM, Lodee MJ, Kalos MD;
PI McNeil PD, Vedvick TS;
XX
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer.
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridization, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELAAACMGGLLLALPPGAASSTVCTGTMKRLPASPETHLIDMKLPHLYOGCGVVGNTL 60
Db 1 MELAAACMGGLLLALPPGAASSTVCTGTMKRLPASPETHLIDMKLPHLYOGCGVVGNTL 60
QY 61 ELTYLPTNASLSFQDIQEVGGYVLIHANOVRQVPLQRLIRVGTQLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFQDIQEVGGYVLIHANOVRQVPLQRLIRVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTTPTVTSASPGRLRELOBSLTTELKSGVLIQNNPOLCTQDTILMDTIFKNNOLA 180
Db 121 DPLNNTTPTVTSASPGRLRELOBSLTTELKSGVLIQNNPOLCTQDTILMDTIFKNNOLA 180
QY 181 LTLIDTNSRACHPCSPCKGSRCKMGSSSEDCQSLTRTVACGACGACCKGLPTDCCHEQC 240
Db 181 LTLIDTNSRACHPCSPCKGSRCKMGSSSEDCQSLTRTVACGACGACCKGLPTDCCHEQC 240
QY 241 AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLHNOEVTAEQDQRCCKRCARVCGLGQYIKANSKFTGIT 360
Db 301 YNYLSTDVSGCTLVCPLHNOEVTAEQDQRCCKRCARVCGLGQYIKANSKFTGIT 360
QY 361 ELEFAGCKKIFGSLAFIPESFDGPASNTAPLOEQVQVETLEITGYLYISAMPDSLP 420
Db 361 IOEFAGCKKIFGSLAFIPESFDGPASNTAPLOEQVQVETLEITGYLYISAMPDSLP 420
QY 421 DLSVFQNIQVIRGRILHNKGAISLTLOGISISMLGRLBELGSLALIHNTHLCPHTV 480
Db 421 DLSVFQNIQVIRGRILHNKGAISLTLOGISISMLGRLBELGSLALIHNTHLCPHTV 480
QY 481 PWDOLFRRPHOALHTANRPEDECVGSGLACHQLCARHGMGPRPTCVNCSQFLRQEC 540
Db 481 PWDOLFRRPHOALHTANRPEDECVGSGLACHQLCARHGMGPRPTCVNCSQFLRQEC 540
QY 541 VEECRVLQGLFREYVNAHCLPCHPECOQPNQSVTCGPEADQCVACAHKDPPEVCYARC 600
Db 541 VEECRVLQGLFREYVNAHCLPCHPECOQPNQSVTCGPEADQCVACAHKDPPEVCYARC 600
QY 601 PSQVPRDLSVMPIMKFPDEBGAOCPCINCHSCVDLDDGCPAEQASLTSIVSVVVG 660
Db 601 PSQVPRDLSVMPIMKFPDEBGAOCPCINCHSCVDLDDGCPAEQASLTSIVSVVVG 660
QY 661 ILVVVGLGVFGLIKRQOKIRKRYTRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILVVVGLGVFGLIKRQOKIRKRYTRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVYLGSGAGFNNFTVSEFWLRP-----KVSASHLEVRRENTSPANKETLDEAVYMA 775
Db 721 RKVYLGSGAGFNNFTVSEFWLRP-----KVSASHLEVRRENTSPANKETLDEAVYMA 775
QY 776 GVGSYVYRLLIGICLTSTVQVLTQLMFYGCLLDVRRNRGLSQDILNMCQIAKMSY 835
Db 776 GVGSYVYRLLIGICLTSTVQVLTQLMFYGCLLDVRRNRGLSQDILNMCQIAKMSY 835
QY 836 LEDVALVRDIAARNVLYKSPNHVKITDFTGLARLLIDETEHYADGQKVIKMALESIL 895
Db 836 LEDVALVRDIAARNVLYKSPNHVKITDFTGLARLLIDETEHYADGQKVIKMALESIL 895
QY 896 RRRPFGSDVWSYGVATWELTFGAKYDGIAPAREIDLEKEBRLPOPICITDVMIM 955
Db 896 RRRPFGSDVWSYGVATWELTFGAKYDGIAPAREIDLEKEBRLPOPICITDVMIM 955
QY 956 VKCWMIDSECRPRRELIVSEFSRMAPDPQRFVVIQNEIDLAPASPLDSTFYRSLDEDDMG 1015
Db 956 VKCWMIDSECRPRRELIVSEFSRMAPDPQRFVVIQNEIDLAPASPLDSTFYRSLDEDDMG 1015
QY 1016 DLVDAEELVPOQGFPCBPDPAPGAGVHRRHSSSTRSGGGLTIGLESSEEARSPSL 1075
Db 1016 DLVDAEELVPOQGFPCBPDPAPGAGVHRRHSSSTRSGGGLTIGLESSEEARSPSL 1075

Db 1016 DLVDAEYIVPQGGFCPPDPAGAGWVHHRSSSTRSGGDLTLGLPSEEPAPRSPL 1075
 QY 1076 APSEGAGSCVFDGDIQMGAKGLQSLPTHDPSPLOQYSEDPVLPSETDGYVAPLTCS 1135
 Db 1076 APSEGAGSCVFDGDIQMGAKGLQSLPTHDPSPLOQYSEDPVLPSETDGYVAPLTCS 1135
 QY 1136 OPEYVNPQVRPQPSPPRGSLPAPRPPGATLERKTLSPKNGVYKDVAFAGAVENPE 1195
 Db 1136 OPEYVNPQVRPQPSPPRGSLPAPRPPGATLERKTLSPKNGVYKDVAFAGAVENPE 1195
 QY 1196 YLTPGGAAPQHPAPAPSPAPNDLYWDQDPERGAAPSTFKGTPTAENPEYGLDVPV 1255
 Db 1196 YLTPGGAAPQHPAPAPSPAPNDLYWDQDPERGAAPSTFKGTPTAENPEYGLDVPV 1255
 RESULT 13
 AAM51143
 ID AAM51143 standard; Protein; 1255 AA.
 AC AAM51143;
 XX
 DT 17-JUN-2002 (first entry)
 XX
 DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
 XX
 KM Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 XX tyrosine kinase; receptor; c-erbB2; gene therapy.
 OS Homo sapiens.
 FH Location/Qualifiers
 FT Domain 1..653 /note= "extracellular domain"
 FT Domain 676..1255 /note= "intracellular domain"
 FT Domain 990..1255 /note= "phosphorylation domain"
 FT
 PN WO200212341-A2.
 XX
 PD 14-FEB-2002.
 PF 03-AUG-2001; 2001WO-US24283.
 XX
 PR 03-AUG-2000; 2000US-0632507.
 XX
 PA (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI; 2002-241743/29.
 DR N-PSDB; ABA92250.
 XX
 PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT extracellular domain response to the protein, has Her-2/neu
 PT phosphorylation domain fused to Her-2/neu intracellular or
 PS
 PS Claim 68; Fig 7, 14pp; English.
 CC The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its deleted fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.
 XX
 SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6602; DB 23; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
 QY 1 MELALCWGILLALPPGASTQVCTGDKRLPASPETHLMLHLVQCGVQGNL 60
 Db 1 MELALCWGILLALPPGASTQVCTGDKRLPASPETHLMLHLVQCGVQGNL 60
 QY 61 ELTYLPTNASLFLQDIEVQGYVLIANNQVPLQRLRIVRGTQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIEVQGYVLIANNQVPLQRLRIVRGTQFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTGASPGGLRELQRLSTELKGVLIQRPQCYQDITLMDIFPKNNQLA 180
 Db 121 DPLNNTTPTGASPGGLRELQRLSTELKGVLIQRPQCYQDITLMDIFPKNNQLA 180
 QY 181 LTLIDTNRSRACHPSCPMCKSRCKGSESEDCOSTRTVCGAGCARCKGPLPTCCHEQC 240
 Db 181 LTLIDTNRSRACHPSCPMCKSRCKGSESEDCOSTRTVCGAGCARCKGPLPTCCHEQC 240
 QY 241 AAGCTGPKHSDCLAFHNHSGICEHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLAFHNHSGICEHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVQPLNQEVTAEQVQRECKSKPARCYGLGMVYKANKFGIT 360
 Db 301 YNYLSTDVSGCTLVQPLNQEVTAEQVQRECKSKPARCYGLGMVYKANKFGIT 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDPGPASTATLQEPQLQVFETLEITGYLYISAMPDLP 420
 Db 361 IQEPAGCKKIFGSLAFLPESFDPGPASTATLQEPQLQVFETLEITGYLYISAMPDLP 420
 QY 421 DLSVFQNLQVIRGLIHNGAYSLLTQGGISMLGRSLRELGSGIALIHNHTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGLIHNGAYSLLTQGGISMLGRSLRELGSGIALIHNHTHLCFVHTV 480
 QY 481 PMDQLFRRPHQALLTANRPEDECGEGGLACHQICARHGCMGPPTQCVCNCSQFLRQEC 540
 Db 481 PMDQLFRRPHQALLTANRPEDECGEGGLACHQICARHGCMGPPTQCVCNCSQFLRQEC 540
 QY 541 VECRYLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCYACAHYKDPFCVARC 600
 Db 541 VECRYLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCYACAHYKDPFCVARC 600
 QY 601 PSGVPRDLSYPMIMKFPBEGACQCPINCHSCYVDLDDCKCPAEORASPLTISAVG 660
 Db 601 PSGVPRDLSYPMIMKFPBEGACQCPINCHSCYVDLDDCKCPAEORASPLTISAVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPQAOIRLKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPQAOIRLKETEL 720
 QY 721 RYKRYLGSABEFNNFTYSPMLRVF----KYASHLVELEENSPRANKETLDEAVYMA 775
 Db 721 RYKRYLGSABEF--TYKGIWIPDGSNVKIPVA-IVLENTSPRANKETLDEAVYMA 775

QY 776 GVGSFVYSRLGICLTSTVQVLTQMPYGCILDHVRENRGRIGSDLLNMCQIAKMSY 835
DB 776 GVGSFVYSRLGICLTSTVQVLTQMPYGCILDHVRENRGRIGSDLLNMCQIAKMSY 835
QY 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALLESIL 895
DB 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALLESIL 895
QY 896 RRRFTHQSDVWSGYVTWELMTFGAKFYDGIIPAREIPDLEKGERLPOPICITIDVYIM 955
DB 896 RRRFTHQSDVWSGYVTWELMTFGAKFYDGIIPAREIPDLEKGERLPOPICITIDVYIM 955
QY 956 VKCMIMIDSECRPFRELVSEFSRMAADPQRFVYIQNEDLGASPLDSTFFYSLLLEDDMG 1015
DB 956 VKCMIMIDSECRPFRELVSEFSRMAADPQRFVYIQNEDLGASPLDSTFFYSLLLEDDMG 1015
QY 1016 DLVDAEEYLVPOGFFCPDPAPGAGMWHRRSSSTSGGSDLTLGEPSEEPAPRSPL 1075
DB 1016 DLVDAEEYLVPOGFFCPDPAPGAGMWHRRSSSTSGGSDLTLGEPSEEPAPRSPL 1075
QY 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPPTVPLPSETDGYVAPLTCSP 1135
DB 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPPTVPLPSETDGYVAPLTCSP 1135
QY 1136 QPEYVNPDPVRPQPSPREGSLPAPAPAGATLERAKTILSPGNKVYKQVFAFGAVENPE 1195
DB 1136 QPEYVNPDPVRPQPSPREGSLPAPAPAGATLERAKTILSPGNKVYKQVFAFGAVENPE 1195
QY 1196 YLTPQGGAAPQHPAPSPAFDNLVYWDQDPERGAAPSTFKGTPTAENBEYLGDLVPV 1255
DB 1196 YLTPQGGAAPQHPAPSPAFDNLVYWDQDPERGAAPSTFKGTPTAENBEYLGDLVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AAU77114;
DT 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
DE Human Her-2/neu polypeptide.
XX
XX Human; Her-2/neu; cytosolic; haematological malignancy; CMU;
KM acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KM chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KM Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
OS
XX
XX MO200213847-A2.
PN
XX
PD 21-FEB-2002.
XX
XX 13-AUG-2001; 2001MO-US25408.
PF
XX
XX 14-AUG-2000; 2000US-0638280.
PR
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
PA
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
XX MPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CMU), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 23; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
QY 1 MELAALCRWGLLALLPFGASTQVCTGIDMKLRLPASPEHLDMRLHYGCGVQGNL 60
DB 1 MELAALCRWGLLALLPFGASTQVCTGIDMKLRLPASPEHLDMRLHYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLDIOEVQGYVLIANQVROYPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGYVLIANQVROYPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLANTPTPTGASPGGLBELQRLSTELIKSGVLIQRNPOLCYQDTILMKDIFHKNOILA 180
DB 121 DPLANTPTPTGASPGGLBELQRLSTELIKSGVLIQRNPOLCYQDTILMKDIFHKNOILA 180
QY 181 LTLIDPNSRACHPSPCKSGRCMGESEDOQSILTRIVCAGGCAKCGPLPTDCHEQC 240
DB 181 LTLIDPNSRACHPSPCKSGRCMGESEDOQSILTRIVCAGGCAKCGPLPTDCHEQC 240
QY 301 YNVLSTDVGSCTLVGCLHNOEYTAEDGQRECKSKPCARVCYGLGMQYIYANSKFICIT 360
DB 301 YNVLSTDVGSCTLVGCLHNOEYTAEDGQRECKSKPCARVCYGLGMHLEBVAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFPDGPASNTAFLQPEQLQVFETLEITGYLISAMPDLP 420
DB 361 IOEPAGCKKIFGSLAFLPESFPDGPASNTAFLQPEQLQVFETLEITGYLISAMPDLP 420
QY 421 DLSVQNLQVIRGILHNGAYSILTQIGISWLGIRSLRELSGLALIHNTHLCEVHTV 480
DB 421 DLSVQNLQVIRGILHNGAYSILTQIGISWLGIRSLRELSGLALIHNTHLCEVHTV 480
QY 481 PMDQLFRNPHQALHTANRPEDECVGEGACIHOUCARGHGCGPPTQCVNSQFLRGQC 540
DB 481 PMDQLFRNPHQALHTANRPEDECVGEGACIHOUCARGHGCGPPTQCVNSQFLRGQC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHBECQPOGNSYTCFGEADQCAAHYKDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHBECQPOGNSYTCFGEADQCAAHYKDPFCVARC 600
QY 601 PSQVPRDLSYMPIMKFPDEBGAQCPICNCHSCVDLDDKCPAEQASPLTSTVSAYVG 660
DB 601 PSQVPRDLSYMPIMKFPDEBGAQCPICNCHSCVDLDDKCPAEQASPLTSTISAYVG 660
QY 661 ILVVVVLGVVFGILIKRQOKIRKTYRRLLOETELVEPLTPSGAMPQAOQMRILKTEEL 720
DB 661 ILVVVVLGVVFGILIKRQOKIRKTYRRLLOETELVEPLTPSGAMPQAOQMRILKTEEL 720
QY 721 RKVKYLGSGAFGFPNNFTYSFWLRVP-----KVASHLVELEBENTSPRANKELDEAYYMA 775
DB 721 RKVKYLGSGAFG---TYKGIWIDGENYKIPVA-IKVARENTSPRANKELDEAYYMA 775
QY 776 GVGSFVYSRLGICLTSTVQVLTQMPYGCILDHVRENRGRIGSDLLNMCQIAKMSY 835
DB 776 GVGSFVYSRLGICLTSTVQVLTQMPYGCILDHVRENRGRIGSDLLNMCQIAKMSY 835
QY 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALLESIL 895
DB 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALLESIL 895

Db	836	LEDVRLVHRDLAARNLVKSPNHVKITDFFGLARLLDIDETEHADGGVPIKMALESIL	895	Matches 1214; Conservative 13; Mismatches 23; Indels 10; Gaps 3;
Qy	896	RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYIM	955	
Db	896	RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYIM	955	
Qy	956	VKCMWIDSECRPRFRELVEFSRMAADPQRFVVIQNEDLGPASPLDSTFFYSLLLEDDMG	1015	
Db	956	VKCMWIDSECRPRFRELVEFSRMAADPQRFVVIQNEDLGPASPLDSTFFYSLLLEDDMG	1015	
Qy	1016	DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL	1075	
Db	1016	DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL	1075	
Qy	1076	APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLOKRYSEDPVLPSEBTDGVAPLTCSP	1135	
Db	1076	APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLOKRYSEDPVLPSEBTDGVAPLTCSP	1135	
Qy	1136	QPEYVQPDVPRPQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVAFAGAVENPE	1195	
Db	1136	QPEYVQPDVPRPQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVAFAGAVENPE	1195	
Qy	1196	YLTPOGGAAPQHPHPAPFAPFADNLYYWDQDPPERGAPPTKGTPTAENPEYLGADVIV	1255	
Db	1196	YLTPOGGAAPQHPHPAPFAPFADNLYYWDQDPPERGAPPTKGTPTAENPEYLGADVIV	1255	
RESULT 15				
AAK39568	AAK39568 standard; Protein; 1433 AA.			
AC	AAK39568;			
DT	07-FEB-1994 (first entry)			
DE	Sequence of c-erbB-2 tumour antigen.			
XX	Tumour antigen; c-erbB-2; glycoprotein.			
XX	Homo sapiens.			
PN	M09316185-A.			
PD	19-AUG-1993.			
PF	05-FEB-1993; 93MO-US01055.			
PR	06-FEB-1992; 92US-0831967.			
PA	(CETUS) CETUS ONCOLOGY CORP.			
PI	(CREA-) CREATIVE BIOMOLECULES INC.			
PI	Houston IL, Huston JS, Oppermann H, Ring DB;			
DR	WPI: 1993-272889/34.			
DR	N-PSDB: AAQ46083.			
XX	New single chain Fv polypeptide binding to C-erbB-2 tumour			
PT	antigen - for imaging or treating breast or ovarian cancer etc.			
XX	Disclosure; pages 48-54; 87pp; English.			
XX	c-erbB-2 refers to a protein antigen expressed on the surface of			
CC	tumour cells such as breast and ovarian tumour cells, which is an			
CC	apptox. 200,000 mol. wt. acidic glycoprotein having an isoelectric			
CC	pt. of about 5.3 (see AAQ46083, AAK39568). X in AAK39583 represents			
CC	the location of a stop codon in AAQ46083.			
XX	Sequence 1433 AA;			
XX	Query March 96.3%; Score 6559; DB 14; Length 1433;			
XX	Best Local Similarity 96.3%; Pred. No. 0;			
Qy	1	MELALCRWGLLALLPPGAASVQVCTGTDKMLPLPASPETHLDMLEHLVQGVQGNL	60	
Db	1	MELALCRWGLLALLPPGAASVQVCTGTDKMLPLPASPETHLDMLEHLVQGVQGNL	60	
Qy	61	ELTYLPTNASISFLDIOEVQGVYLIANQVQVPLQRLIRVGTQCFEDVYALAVDNG	120	
Db	61	ELTYLPTNASISFLDIOEVQGVYLIANQVQVPLQRLIRVGTQCFEDVYALAVDNG	120	
Qy	121	DLNNTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYDITLWMDIFKNNOLA	180	
Db	121	DLNNTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYDITLWMDIFKNNOLA	180	
Qy	181	LTLIDNRSRACHPCSPCKSGRCMGSESEDCQSLTRVACGAGCARCGPLPTCCHEQC	240	
Db	181	LTLIDNRSRACHPCSPCKSGRCMGSESEDCQSLTRVACGAGCARCGPLPTCCHEQC	240	
Qy	241	AACTGPKHSDCLACHTFNHSGICELCPALVYNTDTFESMPNREGRYTFGASCVTACP	300	
Db	241	AACTGPKHSDCLACHTFNHSGICELCPALVYNTDTFESMPNREGRYTFGASCVTACP	300	
Qy	301	YNTLSTDVASCTLVCPILNQEYTAEDGTQCEKSGPCARVCYGLGMQYITANSKFIGIT	360	
Db	301	YNTLSTDVASCTLVCPILNQEYTAEDGTQCEKSGPCARVCYGLGMQYITANSKFIGIT	360	
Qy	361	ELFAGCKKIFGSLAFPESEFPGDASNTAFLQEPOLQVFETLEITGYLISAMPDLP	420	
Db	361	ELFAGCKKIFGSLAFPESEFPGDASNTAFLQEPOLQVFETLEITGYLISAMPDLP	420	
Qy	421	DLSPQNLQVIRGLIHNGAVSLTQSGISLWGLRSLRELSGLALIHNTHLCPHTV	480	
Db	421	DLSPQNLQVIRGLIHNGAVSLTQSGISLWGLRSLRELSGLALIHNTHLCPHTV	480	
Qy	481	PMDFRNFHQAALHTANRPEDECEVGEGLACHQICARHGCMGPGTQCVCNCSQFLRGEC	540	
Db	481	PMDFRNFHQAALHTANRPEDECEVGEGLACHQICARHGCMGPGTQCVCNCSQFLRGEC	540	
Qy	541	VEECRYLQGLPREYVNAARCLPCHBECOPQNSVTCFGEADQVACAHYKDPFPCVARC	600	
Db	541	VEECRYLQGLPREYVNAARCLPCHBECOPQNSVTCFGEADQVACAHYKDPFPCVARC	600	
Qy	601	PSGVKPDLSYMPIMKFPDEGACQCPPLNCTHS CVDLDDKGPABQKASPLTISIISAVG	660	
Db	601	PSGVKPDLSYMPIMKFPDEGACQCPPLNCTHS CVDLDDKGPABQKASPLTISIISAVG	660	
Qy	661	ILLVVVLGVVFGILIKRQCKIRKTYTMRLLQETELVPLTFSGAMPNQAQRIKETEL	720	
Db	661	ILLVVVLGVVFGILIKRQCKIRKTYTMRLLQETELVPLTFSGAMPNQAQRIKETEL	720	
Qy	721	RKYKVLGSGAFGEFNFTUSFMLRVP-----KVSASHLEVLRENTSPKANKELDEAYYMA	775	
Db	721	RKYKVLGSGAFGEFNFTUSFMLRVP-----KVSASHLEVLRENTSPKANKELDEAYYMA	775	
Qy	776	GVGSPYVSRLLGICLTSVQLVTOAMPYGLLDHYRENRGRIGSQDLNMCQIAKAGSY	835	
Db	776	GVGSPYVSRLLGICLTSVQLVTOAMPYGLLDHYRENRGRIGSQDLNMCQIAKAGSY	835	
Qy	836	LEDVRLVHRDLAARNLVKSPNHVKITDFFGLARLLDIDETEHADGGVPIKMALESIL	895	
Db	836	LEDVRLVHRDLAARNLVKSPNHVKITDFFGLARLLDIDETEHADGGVPIKMALESIL	895	
Qy	896	RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYIM	955	
Db	896	RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYIM	955	
Qy	956	VKCMWIDSECRPRFRELVEFSRMAADPQRFVVIQNEDLGPASPLDSTFFYSLLLEDDMG	1015	
Db	956	VKCMWIDSECRPRFRELVEFSRMAADPQRFVVIQNEDLGPASPLDSTFFYSLLLEDDMG	1015	
Qy	1016	DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL	1075	
Db	1016	DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL	1075	

```
Qy      1076 APSEGAGSTVFENDGLGMGAKGLQSLPTHDPSPLOQRYSEDPPTVLPSETDGYVAPLTCSP 1135
Db      1076 APSEGAGSDVFDGDLGMGAKGLQSLPTHDPSPLOQRYSEDPPTVLPSETDGYVAPLTCSP 1135
Qy      1136 QPEYVNOQPDVRFQPPSPREGPLPAARPAQATLERAKTILSPGNQGVKDVFAFGAVENPE 1195
Db      1136 QPEYVNOQPDVRFQPPSPREGPLPAARPAQATLERAKTILSPGNQGVKDVFAFGAVENPE 1195
Qy      1196 YLTPQGAAPQPHPPAFSPAFDNLVYWDQDPFERGAPPSFTKGTPTAENPEYLGLDVPV 1255
Db      1196 YLTPQGAAPQPHPPAFSPAFDNLVYWDQDPFERGAPPSFTKGTPTAENPEYLGLDVPV 1255
```

Search completed: July 22, 2003, 09:17:00
Job time : 41.9319 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-325-339-14

Perfect score: 6852

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPSYGLDVPV 1261

Scoring table: BLOSUM62

Gap: 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	6730	98.2	1255	1	ERB2_HUMAN
2	5931	86.6	1257	1	ERB2_RAT
3	5916.5	86.3	1254	1	ERB2_MOUSE
4	3148	45.9	1210	1	EGFR_HUMAN
5	3124	45.6	1210	1	EGFR_MOUSE
6	2981.5	43.5	1308	1	ERB4_HUMAN
7	2862	43.2	1308	1	ERB4_RAT
8	2688.5	39.2	1167	1	XMRK_XIPMA
9	2439.5	35.6	1342	1	ERB3_HUMAN
10	2365.5	34.5	1339	1	ERB3_RAT
11	1965	28.7	1426	1	EGFR_DROME
12	1749.5	25.5	634	1	ERBB2_MOUSE
13	1703	24.9	604	1	ERBB2_AVIER
14	1630	23.8	540	1	ERBB2_AVIER
15	1600	23.4	703	1	EGFR_CHICK
16	1300	19.0	1323	1	EGFR_MOUSE
17	1142.5	16.7	245	1	ERB2_MOUSE
18	730	10.7	1363	1	ILPR_BRALA
19	696	10.2	1382	1	ILPR_HUMAN
20	693	10.1	1300	1	ILPR_MOUSE
21	693	10.1	1607	1	ILPR_MOUSE
22	691	10.1	1383	1	ILPR_MOUSE
23	690.5	10.1	1372	1	ILPR_MOUSE
24	689	10.1	1477	1	ILPR_MOUSE
25	684	10.0	1297	1	ILPR_MOUSE
26	679.5	9.9	1300	1	ILPR_MOUSE
27	638	9.3	1367	1	ILPR_MOUSE
28	622	9.1	1373	1	ILPR_MOUSE
29	621	9.1	1466	1	ILPR_MOUSE
30	618.5	9.0	1370	1	ILPR_MOUSE
31	617	9.0	1390	1	ILPR_MOUSE
32	602	8.8	987	1	ILPR_MOUSE
33	598.5	8.7	984	1	ILPR_MOUSE

34	596	8.7	976	1	EPB2_HUMAN	P29317	homo sapien
35	590.5	8.6	977	1	EPB2_MOUSE	Q03145	mus musculus
36	588	8.6	972	1	EPB2_XENTIA	Q01736	xenopus lae
37	588	8.6	987	1	EPB2_MOUSE	P54761	mus musculus
38	588	8.6	1114	1	RET_HUMAN	P07949	homo sapien
39	583.5	8.5	984	1	EPB1_RAT	P09759	rattus norv
40	582.5	8.5	985	1	EPB1_XENTIA	Q91571	xenopus lae
41	577.5	8.4	984	1	EPB1_HUMAN	P54762	homo sapien
42	577	8.4	757	1	HT16_HYDAT	P53356	hydatid atten
43	573.5	8.4	1053	1	FAK1_CHICK	Q00944	gallus gall
44	569	8.3	1068	1	FAK1_XENTIA	Q91738	xenopus lae
45	563	8.2	1052	1	FAK1_MOUSE	P34152	mus musculus

ALIGNMENTS

RESULT 1	ERB2_HUMAN	STANDARD	PRT: 1255 AA.
ID	ERB2_HUMAN		
AC	P04626		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	13-AUG-1987 (Rel. 05, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor Protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-cyte cell		
DE	surface receptor HER2) (MLN 19).		
GN	ERBB2 OR HER2 OR NGL OR NEU.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=6618663; PubMed=3003577;		
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,		
RA	Saito T., Toyoshima K.;		
RT	"Similarity of protein encoded by the human c-erbB-2 gene to		
RT	epidermal growth factor receptor.";		
RL	Nature 319:230-234(1986).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=66070181; PubMed=2999974;		
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,		
RA	McGeehan T., Seeburg P.H., Libermann T.A., Schlessinger J.,		
RA	Francake U., Levinson A., Ulrich A.;		
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor		
RT	shares chromosomal location with neu oncogene.";		
RL	Science 230:1132-1139(1985).		
RN	[3]		
RP	SEQUENCE OF 737-1031 FROM N.A.		
RX	MEDLINE=66016725; PubMed=2995967;		
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;		
RA	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the		
RT	c-erbB-1/epidermal growth factor receptor gene and is amplified in a		
RT	human salivary gland adenocarcinoma.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).		
RN	[4]		
RP	VARIANTS VAL-654 AND VAL-655.		
RX	MEDLINE=93194196; PubMed=8095488;		
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;		
RT	"Characterization of a new allele of the human ERBB2 gene by allele-		
RT	specific competition hybridization.";		
RL	Genomics 15:426-429(1993).		
CC	- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-		
CC	ALPHA AND AMPHIREGULIN.		
CC	- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	- SUBCELLULAR LOCATION: Type I membrane protein.		


```

QY 720 ILKETELRKVYLGSGAGVYKGIWIPDGENVKIPVAIVALENTSPKANKELDEAVY 779
DB 714 ILKETELRKVYLGSGAGVYKGIWIPDGENVKIPVAIVALENTSPKANKELDEAVY 773
QY 780 MAGVSPVYSRLIGICLTSTVQLTCLMPYGCILDHRENRGLSGODLNMCMQIAKEM 839
DB 774 MAGVSPVYSRLIGICLTSTVQLTCLMPYGCILDHRENRGLSGODLNMCMQIAKEM 833
QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLIDETEYHADGKVPKIMMALES 899
DB 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLIDETEYHADGKVPKIMMALES 893
QY 900 ILRRFTHQSDVWSYGVTVWMLMTFGAKPYDGIPIAREIPDLLEKGERLPPEPCTIDVTM 959
DB 894 ILRRFTHQSDVWSYGVTVWMLMTFGAKPYDGIPIAREIPDLLEKGERLPPEPCTIDVTM 953
QY 960 IMVKCMWIDSECRPFRELVESEFSRMARDPQRFVYINEDLGPASPLDSTFYSLLEDD 1019
DB 954 IMVKCMWIDSECRPFRELVESEFSRMARDPQRFVYINEDLGPASPLDSTFYSLLEDD 1013
QY 1020 MGLDVADEEYVLPQGGFCPPDPAAGAGVYHRRSSSTRSGGQDLTLGLEPSEEPAPRS 1079
DB 1014 MGLDVADEEYVLPQGGFCPPDPAAGAGVYHRRSSSTRSGGQDLTLGLEPSEEPAPRS 1073
QY 1080 PLAPSEGASDVPGDGLGMAKAGLQSLPTHDPSPGLGRVEDPTVLPSESTDYVAPLTC 1139
DB 1074 PLAPSEGASDVPGDGLGMAKAGLQSLPTHDPSPGLGRVEDPTVLPSESTDYVAPLTC 1133
QY 1140 SPOPEYVNPDPVRQPSPSPREGLPAARPAAGATLERAKTILSPKNGVYKVPFAGGAVEN 1199
DB 1134 SPOPEYVNPDPVRQPSPSPREGLPAARPAAGATLERAKTILSPKNGVYKVPFAGGAVEN 1193
QY 1200 PEYLTPQGGAPQHPHPAPSPADNLYYWDOPPEPAGAPSPFTFGTPTNENBEYGLDV 1259
DB 1194 PEYLTPQGGAPQHPHPAPSPADNLYYWDOPPEPAGAPSPFTFGTPTNENBEYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 2
ERB2_RAT STANDARD; PRT; 1257 AA.
ID ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC AC P06454;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase etdb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
DE
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein."
RL Nature 319:226-230(1986).
RN RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).

```

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD.RES 1141 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD.RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 188 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.6%; Score 5931; DB 1; Length 1257;
 Best Local Similarity 86.7%; Pred. No. 1,2e-309;
 Matches 1096; Conservative 56; Mismatches 102; Indels 10; Gaps 4;

QY 1 MELALLCRWGLLALLPFGAASSTQVCTGTDMLKLPASPETHLDMLRHLYOGGVQVQNL 60
 DB 1 MELAMCRWGLLALLPFGIAGTQVCTGTDMLKLPASPETHLDMLRHLYOGGVQVQNL 60
 QY 61 ELTYLPTNASTLFQDIOEVQVYLIAHQVQVPLQRLRIIVEGTQLFEDNVALAVLUNG 120
 DB 61 ELTYLPANASTLFQDIOEVQVYLIAHQVQVPLQRLRIIVEGTQLFEDNVALAVLUNG 120
 QY 121 DPLNNTTPTVT-GASFGGLRELIQRLSLTEILKGVLIQRNPQLCYQDTILMKDI-FHKXNQL 179
 DB 121 DPQDNVAASTPGRTRPEGLRELIQRLSLTEILKGVLIQRNPQLCYQDNVLMKDVFRKXNQL 180
 QY 180 ALTLIDITNRSRACHCSPMKGKSGRSGSSSEDCCSLRTVCAGGACAKCKPLPTDCCHQ 239
 DB 181 APVDIDITNRSRACHCSPMKGKSGRSGSSSEDCCSLRTVCAGGACAKCKPLPTDCCHQ 240
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPBGRATFGASCCTAC 299
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPBGRATFGASCCTAC 300
 QY 300 PYNVLTVDGSCITVCPALHNOEVTAEQTCRCEKSCPKARVCYGLGMPNFVTSWMLR 359
 DB 301 PYNVLTVDGSCITVCPALHNOEVTAEQTCRCEKSCPKARVCYGLGMPNFVTSWMLR 359
 QY 360 PK-VASHLSEFPAGCKKIFGSLAPLPESFDQDPASNTAPLOPEQLQVFTLEIEITGYLYT 418
 DB 360 PK-VASHLSEFPAGCKKIFGSLAPLPESFDQDPASNTAPLOPEQLQVFTLEIEITGYLYT 418
 QY 419 SAMPDSLPDLISYFQWLYIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 478
 DB 419 SAMPDSLPDLISYFQWLYIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 478
 QY 414 SAMPDSLPDLISYFQWLYIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 473
 DB 414 SAMPDSLPDLISYFQWLYIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 473
 QY 479 HLCFVHTVPMQDLFENPHQALHTANRPEDE-CVGEGLACHQCLCARGHCMGPPTQCVNC 537
 DB 479 HLCFVHTVPMQDLFENPHQALHTANRPEDE-CVGEGLACHQCLCARGHCMGPPTQCVNC 537
 QY 538 SGLFGRQCEVECRVLCGLPREYVNAHCHLCPHEPCOPQNSVTCFGEPAEDQVACAHYK 597
 DB 538 SGLFGRQCEVECRVLCGLPREYVNAHCHLCPHEPCOPQNSVTCFGEPAEDQVACAHYK 597
 QY 534 SHFLRGQCEVECRVLCGLPREYVNAHCHLCPHEPCOPQNSVTCFGEPAEDQVACAHYK 593
 DB 534 SHFLRGQCEVECRVLCGLPREYVNAHCHLCPHEPCOPQNSVTCFGEPAEDQVACAHYK 593
 QY 598 DPPFCVACPCSGVCKDLSYMPIMKFPDEEGACOPCINCTHSCVDLDDKCPRAEGRASP 657
 DB 598 DPPFCVACPCSGVCKDLSYMPIMKFPDEEGACOPCINCTHSCVDLDDKCPRAEGRASP 657
 QY 594 DSSSCVACPCSGVCKDLSYMPIMKFPDEEGACOPCINCTHSCVDLDDKCPRAEGRASP 653
 DB 594 DSSSCVACPCSGVCKDLSYMPIMKFPDEEGACOPCINCTHSCVDLDDKCPRAEGRASP 653
 QY 658 TSIVAVVGLILVVLGVFGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 717
 DB 658 TSIVAVVGLILVVLGVFGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 717
 QY 654 TPIITAVGVLLFLILVAVVGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 713
 DB 654 TPIITAVGVLLFLILVAVVGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 713
 QY 718 MRLKETELRKVKVLGSGAFGVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 777
 DB 718 MRLKETELRKVKVLGSGAFGVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 777

DB 714 MRLKETELRKVKVLGSGAFGVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 773
 QY 778 YVMAGVSPVSRLLGICLTSTVQLVTLQMLPYCCLLDHRNENRGRLGSDLLNMCMQIAK 837
 DB 774 YVMAGVSPVSRLLGICLTSTVQLVTLQMLPYCCLLDHRNENRGRLGSDLLNMCMQIAK 833
 QY 838 GMSYLEDRVLRHDLAARNVLPVSPNPKITDFGLARLLDIDETEHYADGKVPIMKML 897
 DB 834 GMSYLEDRVLRHDLAARNVLPVSPNPKITDFGLARLLDIDETEHYADGKVPIMKML 893
 QY 898 ESILRRPFTHQSDVMSGVTWELMTPGAKPYGIPAREIPDLLEKGERLPQPICTIV 957
 DB 894 ESILRRPFTHQSDVMSGVTWELMTPGAKPYGIPAREIPDLLEKGERLPQPICTIV 953
 QY 958 YMIWVKCMIDSECRPFRELVESEFSMARDPORFVIVIONEDGPASPDPSTFYRSLLED 101
 DB 954 YMIWVKCMIDSECRPFRELVESEFSMARDPORFVIVIONEDGPASPDPSTFYRSLLED 101
 QY 1018 DDMGDLVDAEELVLPQGFPCPDPAAGGVHRRSSSTRSGGDLTLGLEPSEEAR 107
 DB 1014 DDMGDLVDAEELVLPQGFPCPDPAAGGVHRRSSSTRSGGDLTLGLEPSEEAR 107
 QY 1078 RSLPASEGASVDFDGLGMAAKGLQSLPTHPDPLQRYSDPTVPLPSENDGVVAPL 113
 DB 1074 RSLPASEGASVDFDGLGMAAKGLQSLPTHPDPLQRYSDPTVPLPSENDGVVAPL 113
 QY 1138 TCSPOPEYVNPDPVPPSPRREGPLPAARPAATLERAKTISPGKGVKQVFAFGAV 119
 DB 1134 ACSQPEYVNPDPVPPSPRREGPLPAARPAATLERAKTISPGKGVKQVFAFGAV 119
 QY 1198 ENPEYLLPQGAAPPPPAFSAFENLYYNNQDPPERAPSTKGTPTANPEYGL 125
 DB 1194 ENPEYLLPQGAAPPPPAFSAFENLYYNNQDPPERAPSTKGTPTANPEYGL 125
 QY 1258 DVPV 1261
 DB 1254 DVPV 1257

RESULT 3
 ERB2 MESAU STANDARD; PRT; 1254 AA.
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase etrb-2 precursor (EC 2.7.1.112)
 DE (p15erbB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Cricetinae;
 CC Mesocricetus.
 OC NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T, Ushijima T, Ishizaka Y, Nagao M, Arai M,
 RA Yamazaki Y, Ishikawa T;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY BGF, TGF-
 ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERB2 RECEPTORS
 (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16295; EAA03601.1; -.
 CC HSSP: P11362; IFGK.
 CC Interpro: IPR000494; EGFR_L_domain.
 CC Interpro: IPR000719; Euk_Pkinase.
 CC Interpro: IPR001245; Furin-like.
 CC Interpro: IPR001245; Tyr_Pkinase.
 CC Interpro: IPR004019; YLP_Totif.
 CC Pfam: PF00069; Pkinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF02757; YLP; 2.
 CC Prodom: PD000001; Euk_pkinase; 1.
 CC SMART: SM00261; FU; 3_pkinase; 1.
 CC SMART: SM00219; Tyrc; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferrase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC KW Proto-oncogene; Disease mutation.
 CC FT SIGNAL 1 21 POTENTIAL.
 CC FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 CC FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 653 675 POTENTIAL.
 CC FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 158 368 CYS-RICH.
 CC FT DOMAIN 472 644 CYS-RICH.
 CC FT DOMAIN 720 987 PROTEIN KINASE.
 CC FT NP_BIND 726 753 ATP (BY SIMILARITY).
 CC FT BINDING 753 753 ATP (BY SIMILARITY).
 CC FT ACT_SITE 845 845 BY SIMILARITY.
 CC FT DISULFID 195 204 BY SIMILARITY.
 CC FT DISULFID 199 212 BY SIMILARITY.
 CC FT DISULFID 236 244 BY SIMILARITY.
 CC FT DISULFID 240 252 BY SIMILARITY.
 CC FT DISULFID 255 264 BY SIMILARITY.
 CC FT DISULFID 268 295 BY SIMILARITY.
 CC FT DISULFID 299 311 BY SIMILARITY.
 CC FT DISULFID 315 331 BY SIMILARITY.
 CC FT DISULFID 334 338 BY SIMILARITY.
 CC FT DISULFID 511 520 BY SIMILARITY.
 CC FT DISULFID 515 528 BY SIMILARITY.
 CC FT DISULFID 531 540 BY SIMILARITY.
 CC FT DISULFID 544 560 BY SIMILARITY.
 CC FT DISULFID 563 576 BY SIMILARITY.
 CC FT DISULFID 567 584 BY SIMILARITY.
 CC FT DISULFID 587 596 BY SIMILARITY.
 CC FT DISULFID 600 623 BY SIMILARITY.
 CC FT DISULFID 626 634 BY SIMILARITY.
 CC FT DISULFID 630 642 BY SIMILARITY.
 CC FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 CC MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 CC FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
 CC FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F25B1 CRG64;

Query Match 86.3%; Score 5916.5; DB 1; Length 1254;
 Best Local Similarity 86.3%; Pred. No. 7,3e-109;
 Matches 1089; Conservative 66; Mismatches 98; Indels 9; Gaps 3

QY 1 MELALCRWGLLALLPPGAASVCTGTDMLKRLPASPETHLDMRLHYGQVQVGNL 60
 DB 1 MELAMCWMGLLALLSPGASGTCVCTGDMKRLPASPETHLDIVRLHYGQVQVGNL 60
 QY 61 ETTYPTNASLSPDIOEVOGYVLIANVOYVPLRLRLIVRGTOJFEDNYALAVDNG 120
 DB 61 ETTYPTNASLSPDIOEVOGYVLIANVOYVPLRLRLIVRGTOJFEDNYALAVDNR 120
 QY 121 DELNNTPTVGTASPGGLREILQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFKKNOLA 180
 DB 121 DELNNTPTVGTASPGGLREILQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFKKNOLA 180
 QY 121 DELNNTPTVGTASPGGLREILQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFKKNOLA 180
 DB 121 DELNNTPTVGTASPGGLREILQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFKKNOLA 180
 QY 181 LTLIDTNSRACHPCSPMKCSRWGSSSEDCQSLTFTVACAGCARCKGLPDDCHEQC 240
 DB 181 LTLIDTNSRACHPCSPMKCSRWGSSSEDCQSLTFTVACAGCARCKGLPDDCHEQC 240
 QY 241 AAGCTGPHSDCLACIHNHSGICELCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPHSDCLACIHNHSGICELCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLYCPHLNQSVTAEADGTORCEKCSKPCARVCTGLGFMNFVTSFWLRVP 360
 DB 301 YNYLSTDVGSCTLYCPHLNQSVTAEADGTORCEKCSKPCARVCTGLGFMNFVTSFWLRVP 360
 QY 361 K-VSASHLEEFAGCKPIGSLAFIPESPDGPANTPLOPEQLQVETLEITGYLYIS 410
 DB 361 K-VSASHLEEFAGCKPIGSLAFIPESPDGPANTPLOPEQLQVETLEITGYLYIS 410
 QY 420 AMPDLSPLSVFONLOVIRGSLIHNGAVSLTLOGISWGLRSLRELSGLALIHNNTH 470
 DB 420 AMPDLSPLSVFONLOVIRGSLIHNGAVSLTLOGISWGLRSLRELSGLALIHNNTH 470
 QY 480 LCPVNTVWMDLFRPHALLHTANRPEDVEVGGLACHQLCARGHGWGPGPTQVNCV 530
 DB 480 LCPVNTVWMDLFRPHALLHTANRPEDVEVGGLACHQLCARGHGWGPGPTQVNCV 530
 QY 540 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECOPONGSVTCGPPADOCVACAHKDP 590
 DB 540 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECOPONGSVTCGPPADOCVACAHKDP 590
 QY 594 PFCVAPRCGSKVYKPDLSYMPKFPDEBGAQCPPICTHSCVDLDDKGCAPDEQASPLTS 650
 DB 594 PFCVAPRCGSKVYKPDLSYMPKFPDEBGAQCPPICTHSCVDLDDKGCAPDEQASPLTS 650
 QY 660 IYSAVAVGLLVVGVGVGILIKRROOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 710
 DB 660 IYSAVAVGLLVVGVGVGILIKRROOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 710
 QY 720 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIXYLRENTSPKANKELDEAYV 770
 DB 720 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIXYLRENTSPKANKELDEAYV 770
 QY 780 MAGVSPFVSRLLGLCLTSTVQVLTQMLPYGCLLDHRENRGRIGSODLLNWCQIAKGM 830
 DB 780 MAGVSPFVSRLLGLCLTSTVQVLTQMLPYGCLLDHRENRGRIGSODLLNWCQIAKGM 830
 QY 840 SYLEAVRLVHRDLAARNVLVSPNHVKITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 DB 840 SYLEAVRLVHRDLAARNVLVSPNHVKITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 QY 900 ILRRFTTHQSQSWYSGVTVMELMTFGAKPYDGLIARBEIPDLLEGEGLPOPICTIDIVM 950
 DB 900 ILRRFTTHQSQSWYSGVTVMELMTFGAKPYDGLIARBEIPDLLEGEGLPOPICTIDIVM 950
 QY 960 IIVKCMWIDSECRPFRELVSEFMRMADPORFVIVINEDIGSPASPLDSFFYSLEDDD 1010
 DB 960 IIVKCMWIDSECRPFRELVSEFMRMADPORFVIVINEDIGSPASPLDSFFYSLEDDD 1010

RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RX PHOSPHORYLATION.
 RA MEDLINE=89278137; PubMed=2543678;
 RA Maxolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.W.,
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-glycosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TBGR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB51063.1; -;
 DR EMBL; U48722; AAC50802.1; -;

DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;

Query Match 45.9%; Score 3148; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 6e-161;
 Matches 631; Conservative 177; Mismatches 351; Indels 112; Gaps 2;

11 LLLALPPGAA--STOVCTGDMLELRLPASPEHLLDMLRHLYGGCCVGVGNSLETLYPTN 68
 14 LLAALCPRLALEKRYKQVGGTSNKLQGTGFEDFSLQKFFNNCEVVLGNLEITTVQRN 73
 69 ASLSFLQDICEVGVYLIANQVQVPLQRLIRYVQTQLEFEDNYALAVLDNGPLNNTTP 124
 74 YDLSLFKTIQVAGVLIALTVERIPLENQIRGNMYEENSVALAVLSYVD----- 124
 129 VTGASPGRLRLQLRSLTEILKGVVLQRPQLCYOTILMKDFHKNQALALITDTR 181
 127 --ANKTELKELPMKNLQETILHGAVRESNNPALCNVESIQMRDIVSSDFLSNMSMDQNH 181
 189 SRACPCSPMKGSRGCMGSESDCQLTRTVACAGCA-RCKGPLPTDCHHQCAAGCTGP 241
 184 LGSQCKDPCSPGNSCMGAGBENQKLTIKIQAQCSGRGCRGKSPDCCNQCAGCTGP 241
 248 KHSQCLALFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACPYNYLSTD 301
 244 RESDVLCKRFDEATCKDCTCPPLMLNPTTYQMDVNPBGKYSFGATCVKCPANVYVTD 301
 308 VGSCTLVCLPILNNOETVADGTQRCCKSKPCARCYGLGMFNNTVFMLRVPRVVASHL 361
 304 HSCVTRACGADSYEH-EDGVKCKKCEGCRKRVKNGIGI-GEFKDYL----SINKATNI 356
 368 EFAPACKTIFGSLAFLPESPDPDASNTAPLQBPQLQVFETLEITGYLYISAWPSLDP 421
 357 KHFKNCTISGDLHLPLVAFRGSDFTHPLDPEGLDLTKVKEITGFLLIQAMPERTD 414
 428 LSVFQNLQVTRGLIHNGAVYSLTQGLISLGLRSRLSGSLALHHNTHLCFVHTVP 481
 417 LHAFFNLLEIIGRTKQHOQFSLAVASLNTISLGRSLKESIDGVIITISGNKNLCYANTIN 474
 488 WDLFRNHOALLHTANRPEDECEVGEGLACQHLQCARHGCMGPGTQVCNCSQPLRQECV 541
 477 WKSLFGTSGQTKTIISNRGENSCATQOVCHALCSPEGCMGPEPRDVCNRNVRGECV 534
 548 EECRYLQCLPEEYVNAHRCLEPHEPCOPQNSVTCSFPEADQCVACAHYDPPCVARCP 601
 537 DKCNLLEGEPRPEFENSECICQHEPCLPQAMNITCTGRGPDNCIQCAHYIDGPHCVTCP 594
 608 SGVKRDLSYMBILWKPDEGACQCPINCTHSCVDLDDKQCPAQRASPLTSYSAVVG- 664
 597 AGWGEENNTL-VMKYADGHVCHLCHPCTYGTCTGPGLEGCTPNGPRI--STATGVGA 651
 667 --ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVLEPLTPSGAMPQAQRILKET 724
 654 LLLLLVVALGIG---LFMRRHRIYKXTRLRLDERELVEPLTPSGAPNALRLIKET 710
 725 ELRKXVLSGAFGVYVGIITPGENVVKIPVAKVLRNTSPRANKETIDDEAVYVAGVG 784
 711 EFKKIVLGSAGFGVYVKGWIMIPGEKVKIPVALIKELREATSPRANKETIDDEAVYVAVSD 770
 785 SPVYSRLIGICLTSTVQLTOLMYPYGCILHVBENRRLSSODLLNMCQIAGMSYLED 844
 771 NPHVRRLIGICLTSTVQILTOLMFGCLLDVRRHKNKISSOYLNMVQIAGMSYLED 830
 845 VLVHRLDLAANNVLKSNHVKITDPGLARLLIDEPFYHADGGKVPKIMMALESILRRR 904
 831 RLVHRLDLAANNVLKTPQHKIIDFGLAKLIGAEKEKYAEGKVPKIMMALESILHRI 890
 905 FTHQSDVNSYGVATWELMTFGAKPYDGIIPAREPDLLEKEKRLPOPPICTIDVYMMIMVK 964
 891 YTHQSDVNSYGVATWELMTFGSKPYDGIIPASEISITLKEKRLPOPPICTIDVYMMIMVK 950

```

QY 965 WMIDSECRPRELVESEFSRMARDPQRFVVIQ-NEDLGAPASPLDSTFYRSLHEDDMDGL 1023
DB 951 WMIDADSRKRELLIEFSKARADQRLVIGDGERMLPPTSTSNRYRALMDEEDMDV 1010
QY 1024 VDAEYLVPQCGFCPPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEARSLAP 1083
DB 1011 VDAEYLVPQCGFF-----SSPSTSRPLLS 1036
QY 1084 SEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSET--DGYVAPLTCSP 1141
DB 1037 SLSATSN--NSTVACIDRNGLOSCPIKEDSFLOKRYSDPTALTLEDSDTL-----P 1088
QY 1142 QPEYVNPDPVRPQPSREBPLPAAPAGATLEAKTLSPGKNGVXDVAFGAVENPE 1201
DB 1089 VPEYINQ--SVPRKPAQSVQNPVYHNQPLNP-----APSRDEHYQD--PHSTAVGNPE 1137
QY 1202 YL-TPOGGAAPQHPHPPAFSPFDNLYWMDQ-----DP-----PERGAPPTFK 1244
DB 1138 YLNTVQ-----PTCVNSTFDSPPAHMAQKSHQSLDNPYQDFFPKAKNGIFK 1188
QY 1245 GTPTAENPEYL 1255
DB 1189 GS-TAENAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KCGR) gene."
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation."
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Barp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase."
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site."
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Bissinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP130 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce
CC or send an email to license@isb.ch).
CC -----
DR EMBL, X78987; CAA55587.1; -
DR EMBL, U03425; AAA17899.1; -
DR EMBL, X59688; CAA42219.1; -
DR EMBL, L06864; AAA53029.1; -
DR EMBL, Z12608; CAA78249.1; -
DR HSSP, P11362; 1FGK.
DR MGD, MGI:95294; Egfr.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam, PF00069; Pkinase; 1.
DR Pfam, PF00757; Furin-like; 1.
DR Pfam, PF01030; Recep_L domain; 2.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART, SM00261; FU; 3.
DR SMART, SM00219; Tyrc; 1.
DR PROSITE, PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE, PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE, PS50011; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT DOMAIN 714 981
FT NP_BIND 720 728
FT BINDING 747 747
FT ACT_SITE 839 839
FT DISULFID 190 199
FT DISULFID 194 207
FT DISULFID 215 223
FT DISULFID 219 231
FT DISULFID 232 240
FT DISULFID 236 248
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333

```

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 582 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1112 1112 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690820D46DF2D2F5 CRC64;

```

Query Match 45.6%; Score 3124; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 1,1e-159;
 Matches 635; Conservative 172; Mismatches 353; Indels 120; Gaps 26;

```

QY 11 LTLALLPPGAA--STOVCTGTDMLKRLPASBETHLDMRLHYQCCVYQGNLEETVLPNT 68
DB 14 LTLALCAAGALAEKXVCOGTSNRLTQGTEDHFLSLQGNMNCVELVGNLEETVYQGR 73
QY 69 ASLSFLQDIQEVQGVYLIANQVRQVPLQRLIRYRGQLFEDNYALAVLDNGPLNNTP 128
DB 74 YDLFLKTIQEVAGVYLIANTVERIPLENQIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLREIQLRSLTEILKGVLYIQRNPQLCYODPTILMKDI---FHKNQALTLI 184
DB 125 -YGTNRITGLRELPKRNLOEILIGAVRFSNNPILCNMTIQRDVIQVVFVSNMSMDL--- 180
QY 185 DTNRSRACHPSPCKGSRGWESSESDCQLTRTVACAGCA-RCKGPLPTDCCHEQCAAG 243
DB 181 -QSHPSSCPCKDPSGNSCWGGGEENCQKLTIKIICAQCSHRCRGRSPDCCNCCAG 239
QY 244 CTGPKGSDCLALHNHSGICELHCPALVYNTDFTFSMNPBERITYFGASCYTACPRNY 303
DB 240 CTGPRSPDLVQCQFODEATCKDTPMLNPTTYMDVNPBGKYSFGATCVKCKPRNY 299
QY 304 LSTDVSGCTLVCPHNOEVTAEQTORCEKCSKPCASVYGLGM--FNNFTVAFMLRVK 361
DB 300 VYTDHSGCVRAQCGPRYEV--EEDGIRCKCKDDGCRVAVNGIGIGERKD-TLS----- 350
QY 362 VSASHLEEFACCKIFGSLALPSSFGDDPSANTAPLQPELOVFEETLEITVYISAW 421
DB 351 INATNIKHFYCTAISGDLHLTPAFKGDSEFTTRPPDPELEITLKTKVKEITFLLIQAM 410
QY 422 PDSLPLDSVFNLOVIRGRILHNGAYSLTQGLGISWGLRSLRELGSLALIHNTHC 481
DB 411 PDNMTDLHAFENLEIRGRITKHQHFSLAVGLNITSLGRSLKEISDGVITISGRNLC 470
QY 482 FVHTVPMQDLFRNPHQALLHTANFPEDECVGEGIALCHQLCARHGCWGPFGFTQVNC 541
DB 471 YANTINMKKLFGTNPQKTKIMNRAEKDCAVNVHCNPLCSSGSGCGPEPRDCVSCONS 530
QY 542 RQGECECECVLQGLPREYVVARHCLPCHPECCQPNQNSVTCFPGEDADQCVACAHYNDPFF 601

```

```

DB 531 RGRECEKCNILLEGPREFVENSECIQCHEPCLPQAMNITCTGGRPNQICAHYIDGPH 594
QY 602 CVARPSGVKDDLSWPIWKFPDEBAGCQCPINICTHS CYVDLDKGRPARQASPLSIV 66:
DB 591 CVKTPAGIKENNNTL-VWKYADANNVCHLCHACTGACGPGIQGCEWPSGKIPISA 64:
QY 662 SAVVGLLVVVLGVVFGI-LIKRQOKIRKTYRRLIQETELVEPLTPSGAMPQAQMRI 72:
DB 650 TGIYGLGLFIV-VALGILGFMRRRHVIRKTLARLLQERELVEPLTPSGAPRQALRI 70:
QY 721 LKETELRKVYLGGAGFTYTKGIWIDGENVKIPVAIKVIRENTSPKANKELIDEXVM 78:
DB 709 LKETEFKXIKVLGSGAGFTYTKGIMPEGEVKVIPAIKELRENTSPKANKELIDEXVM 76:
QY 761 AGVSPYVSRLLGCLTSTVOLQMLPYGCLDHDVENGRGLDOLLMCMQIAGMS 84:
DB 769 ASVDNPHVCRLLGICLSTVOLITQMLPYGCLDHYREHKDNISQYLLMNCVOIAGMN 82:
QY 841 YLEDVRLVHRDLAARNVLYKSPNNVKTIDFGLARLIDIDETVHADGKVPIMMALESI 90:
DB 829 YLEDRLVHRDLAARNVLYKTPQVKTIDFGLAKLIGAEKEVYHAEQKVPIMMALESI 88:
QY 901 LRRRTQSDVWSYGVYVWELMTGAPRYDGIAPREIPDLLEKGERLPQPPICITDVMY 96:
DB 889 LHRITQSDVWSYGVYVWELMTGSKPYDGIAPASDISILEKGERLPQPPICITDVMY 94:
QY 961 MYKCMWIDSECRPRELVESEFSHMARDPQRFVYIQ-NEDLGSPASPLDSFVYSILEDD 101
DB 949 MYKCMWIDADSRPFRELILEFSQAARDPQRYLYIQDDEMHLPSPIDSIFYALMDEED 100
QY 1020 MGDLVDAEYVLPQGGFCPPDPAPGAGMYHHRRSSSTRSGGDLTLGLEPSEEARPS 107
DB 1009 MEDVVDADAEYVLPQGGF-----NSPST-----SRT 103
QY 1080 PLAPSEAGSDVPFGDQGMGAKGLSLPHDPSPLRYSSEDPVLPSECT--DGYVAPL 113
DB 1035 PLSSLSATSN-----NSTVACINRNSCRKEDAFELRYSSDPFGAVENTIDAFU--- 106
QY 1138 TCSPOPEYVNPQVPRPPSPREGPLPAPAPAGATLERAKTISPGKXGVVYDFAFGAV 115
DB 1088 --FVPEYVNG-SVPKXPBAGVQVPPVYHNPRLP-----APGRDLQHN--PSNAY 113
QY 1138 ENPEYL-TPQGAAPQCHPPAPSPADNLYYDQ-----DP-----PEKCAP 124
DB 1134 GNPEYLNTPAQ-----PTCLSSGNSPALWIOGSHQMSLDNPYQDFFPKETKPN 118
QY 1241 STFKGTPAENPEYLGDPV 1260
DB 1185 GIFKG-PTAENAEYLRVAP 1203

```

RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 ID ERB4_HUMAN
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=91189574; PubMed=8383326;
 RA Ploeman G.D., Culicseou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 RA Foy L., Neubauer M.G., Shoyab M.,
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 tissue distribution and differential processing in response to
 RT photol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07866; AAB59446.1; -.
 DR HSSP: P13622; 1FCK.
 DR Gene: HGNC:3432; ERBB4.
 DR MIM: 600543; -.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 4.
 DR SMART: SM00219; TyrKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308 POTENTIAL.
 FT DOMAIN 26 651 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT TRANSMEM 652 675 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 676 1308 POTENTIAL.
 FT DOMAIN 186 334 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.

FT NP BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT SITE 843 843
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NGPSTHCITYPMWGHSTLQHA -> IGSSIEDCIGLMD
 FT (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80965D88761 CRC64;
 Query Match 43.5%; Score 2981.5; DR 1; Length 1308;
 Best Local Similarity 45.2%; Pred. No. 5.1e-152;
 Matches 611; Conservative 187; Mismatches 375; Indels 179; Gaps 29
 QY 9 WGLIALLPQAA-----STGYCTGTDKRLPASPETHLDMLRHLXGCGVQGNLELT 64
 DB 8 WWWVSLVLAAGTVQPSDSQSCVCACTEKKSLSDLEQYBALRKYVENCVMGNLEITS 67
 QY 65 LPTNASLSFLDDIOEVQGYVLIANOVROVPLORLRIVRGTOLEFEDNYALAVLDNGDPLN 124
 DB 68 IEHNNDLSFLRSVREYGVVVALNORRYVPLELRIRIRTKLYEDVYALAIPLNRRKDG 127
 QY 125 NTPVPTGASPGALRELOLRSLTEILKGVLIORPOLCYODTILMKDIFKNNQALTLTI 184
 DB 128 NF-----GLDELGKNTLTLNGVYVDQDKFLCYADTIHMQDIVNPPNSLITLV 178
 QY 165 DTRSRACPCSPCKSGRCWGESSEDCGLTRTVCAAGC-ARCKGFLPTDCHQCCAG 243
 DB 179 STNGSSGCGRHKSCCTG-RCWGPTENHCQTLTRTVCAECCDGCYGGVYSDCCRCACAG 237
 QY 244 CTGPGHSPCLCLHPNHSIGIELHCPALVLYNDDTFPSMNPBGARYFGSCVTAAPVNY 303
 DB 238 CSGPDTDCFCMKNFNSGACVTCQPPQTFYNNPTTFLENFAKTIYGAFCYKCKPPHF 297
 QY 304 LSTDVGSCTLVCPHANOEVTAEDGTQRCCKSKRCARVCGYGLMFMNFTVSPFLRVPKYS 363
 DB 298 V-VDSSSCVRAKCPSSKREV-EENGSIKKCKECTDLCPRACDGI-----TGLSMSAQTV 349
 QY 364 ASHLEEFQCKKIRGSLAFLPESDGDPAANTAPLQEOQLQVETTELIGYVYIAMD 423

```

Db 350 SSNDKFINCKINGNLI FLVTG:HGDPYNAIEAIDEKUNVTRVREITGILNIGSWBP 409
Qy 424 SLPLSVFQNLQVIRRIHNGAYSLTLQGLGISMWLGJASRLREISGLALIHNTHLCEV 483
Db 410 NMTDFSVNSLVTTIGRVLYSGSLILLIKOOGITSLQPSQKESLAGNITYITNSNLCTY 469
Qy 484 HTVPMDOJFBNPHQALHTANRPDECEVSGSLACHQUCANGHCWGPQPGTCVNCSPFLRG 543
Db 470 HTIMWTLFTSTINORIVIRDNKKAENCTAEGMVCNHLCSDDGCGPDPCLSCRFBSRG 529
Qy 544 QCVBECRVLQGLREYVNAHNCI:PCHECOOP -ONGSTGEGPADOCVACXKXDPBPC 602
Db 530 RICHSCNLYGFERREFENGSI:CECDBQCKXMEDGLTCHGEPDNYCKSHFKGQPNK 589
Qy 603 VAPCSGVKPDLSYMPWKFPDEBACQCPINCTHSCVLDJDXGC-----PAE 651
Db 590 VEKCPDGLQGANSF--IFKYADPDRCHPCHPNCTQCGNPTSHDCLYPMWGHSTLPHQ 647
Qy 652 QASPLTIVSANY-GILLVYVGLVFGILIKRQCKIKRYTMRLLQETELVEPLTPSG 710
Db 648 AR-TPL--IAAGVIGGLFVLIVGLTAVVYRRKSIX-KRRALRFL-ETELVEPLTPSG 702
Qy 711 AMPNQAQRILKETELKRVKVLGSGAFGVYKGIWIPDGENVXIPVAKVLRNTPSKAN 770
Db 703 TAPNQAQRILKETELKRVKVLGSGAFGVYKGIWVEGETVXIPVAKITLNTTGPKAN 762
Qy 771 KEILDEAVVAGVSGPYVSRLLGICLTSTVQVLTQMPYQGLLDHVENRGRIGSDDLIN 830
Db 763 VEFMDEALIMASMPHLVRLGLCLSPITQLVQMPHQCILEVYEHKDNIGSOLLIN 822
Qy 831 WMOJAKGMSYLEDRVLVHRDLAARNLVKSPNVKXTDGLRLDIDETEHAGGKV 890
Db 823 WCVQJAKGMYLLEERVLVHRDLAARNLVKSPNVKXTDGLRLDIDETEHAGGKV 882
Qy 891 PIKMMALLESILRRFTHQSDVWSYGVTVWELMTFGANPYDGIAPAREIPDLLEKGERLPQ 950
Db 883 PIKMMALLECIIHYRFTHQSDVWSYGVTVWELMTFGKXPYDGIPTREIPDLLEKGERLPQ 942
Qy 951 PICTIDVYMWKCMMDIDSECRFRFRFVSEFSMADPQRFVYVIONED-LGRASLJST 1009
Db 943 PICTIDVYMWKCMMDIDSECRFRFRFVSEFSMADPQRFVYVIONED-LGRASLJST 1002
Qy 1010 FYRSLLEDDMDGLVDAAEYLVPOQGFCCDPAPAGAGMWHHRSSSTPSGGDLTLGL 1069
Db 1003 FFQNLDEDELEDMDAAEYLV-CAFNIPPP-----ITSAARIDSNNS-----EIGH 1050
Qy 1070 EPSEEBAPRS-----PLAP-SEGAQSCVFPDGLGMA 1100
Db 1051 SPAPAYTPMSGNQFVYRDGGAAGVSVYRAPRTSTIPAPVAOGATAIFPDSCCNCT 1110
Qy 1101 AKGLQSLPTHDPSPLQRYSEDPVPLPS-----ETDGVVAPLTCSPQPEYVNOQDVP 1153
Db 1111 LRKVAHVHVQEDSSTQYSAPTYFAPERSPRGSLDEBGMTMRDKPKKEYINPVE--- 1167
Qy 1154 QPSPRREGPLPAADPAGATLERAKTSLPGKGVVYKDFAFGAVENPEYLTPOGGAAPQ 1213
Db 1168 -----ENPFVSR-----KXGDIQ-----ALDNPEYHNASNG----- 1194
Qy 1214 HPPA-----FSPADNLVYWDQPPPERGA- 1238
Db 1195 -PFKAEDRYNEPLYLNTPANTLGKAEYLNKONLISMPKAKKAFDNDPWNHSLPRSTL 1253
Qy 1239 -PSTFKGTPT-----AENPEYL 1255
Db 1254 QHPDYLGQYSTKYFYKONGRIRPIVAENPEYL 1265

```

```

RESULT 7
ERR4_RAT
ID ERR4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni W.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269 (1998).
RN (2)
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704 (1991).
RN (3)
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Froehner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; IFKG.
CC InterPro: IPR000494; EGF_R_L domain.
CC InterPro: IPR000719; Euk_Dkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_Dkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; Pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP_2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.

```

DR SMART: SM00261; FU: 4.
DR SMART: SM00219; TYRKC: 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM_1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PVT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;

Query Match 43.2%; Score 2962; DB 1; Length 1308;
Best Local Similarity 45.0%; Pred. No. 5,6e-15;
Matches 609; Conservative 194; Mismatches 384; Indels 166; Gaps 29;

QY 1 MEIA-ALCRWGLL--ALLPGAASVCTGTDMKRLPASPEHLMDLRLHLYGSCQVQ 57
DB 1 MKLATGLWVGSLVAARTVCPASQASQACGTENKLSLSDLEQYBALRKYENCCEVM 60

QY 58 GNLELTLYLPNTASLSLSDIOGVGVYLIANOVROVLORLRYRGQOLFEDNYALAVL 117
DB 61 GNLELTISIEHNRDLSLSIREVTGYVALNQPFLPLLENLRLIRGKLYEDSYALAI 120

QY 118 DNGDPLNNTPTVTGASPGGLRELQRLSTLEILKGVLIQRPOLCYQDPTLMKDI 177
DB 121 LNVKKGNF-----GLQELGLKNLTIELINGVYVDQNKFLCYADTIHWQDIV 171

QY 178 QLATLIDITNRBRACHPSCPMCKSRGCMGSESSDOSTRTVAGGC-ARCKPLPTDCC 236
DB 172 PSNMTLVSTIGSSCGGCHKSCIG-RMGSTENHQILRTVCAEQDRCVFPYSDCC 230

QY 237 HEQCAAGCTGPKHSDCLACHFNHSGICEJHCPALVTYNTDTESMNPREGRTFGASCV 296
DB 231 HRECAAGSGGPKDIDCEACMNFNDSGACVYQCQTFVYNTTQLEHNPFAKYTAFAVC 290

QY 297 TACRYNLTSTVGSCITVCPHLNQEVAEDGTORCEKSPCARCVGGLGMFNNFTVFM 356
DB 291 KCPENFV-VDSSCVACPSKMEV-EEGIMCKRCTIDCKRACDGI-----TGLS 342

QY 357 LRVKVSASHLEEFAGCKKIFGSLAPLPESPDPAASNTAPLOPEOLQVETLEITGYL 416
DB 343 MSAQVDSNIDKFINCTKINGNLIPLVTGIHGPYVAIDAIDPEKLVRTVREITGFL 402

QY 417 YISAPDSLPPLSFQVLYIRGILHNGAVSLTGLGISMGLSLRSLGSLALHH 476
DB 403 NIQWPMPTDVSFNSLVTTGGVLSGLLILKQGITSLQFOSLKEISAGNIYTD 462

QY 477 NTHLCFVHTVPMQLFNNPQALHTANPEDECVGEGLAQHOLCARGHCMGPGPTCVN 536
DB 463 NSNLCTYHTITWTLTFTVNGRIYIRNRAENCTAGMWCNHLCSMDGCMGPGDCLS 522

QY 537 CSQPLRGQCEYEBRVLYQGLPREVYNAHCLPCHPBCQP-QNSVYTFGFEBAQVCAAH 595
DB 523 CRPSRGKICIESCNLYDGEREFENSGIIVECDQCEKREDGLTCHGPGPNCTKSH 582

QY 596 YKDPFCVACRPSGVKEDLSYMPMKPEDEGACQCPRICTSHSCVLDLDDKC----- 646
DB 583 FKDGNCYKPEVDVLOGANSF--IFKADDDRECHPHNCTQCGNPTSHDCTIYWTG 640

QY 649 ----PAQRASPLTSIYSAVY-GILVVVVGAVFGILIKRQOKIRRYTRRLLOETELV 703
DB 641 HSTLPQHR-TPL-IAAGVIGGLFIVLIMALTFAVYVRKRSIK-KRALRRL-ETELV 695

QY 704 EPLTSGAMPQAOQMLIKETELKXVVLSSGAGTYGKIMIDGENNYKIPVAKVIRE 763
DB 696 EPLTPSGAPQAOQRLIKETELRVVLSGAGTYGKIMIDGENNYKIPVAKVIRE 755

QY 764 NTSPEANKEIDEAVYVAGVSPVYSRLGICLTSTVQVLTQVLTQMPYGLDHYENRGL 823
DB 756 TTGPANVENFDEALIMASVDRPLVLLGVCCSPITQVLTQMPHCCLEHYEHKDN 815

QY 824 GSQDLNMCQIAGMSTLEVRVYHNDLAARNVLYVSPHVKITDGLARLIDIDTEY 883
DB 816 GSQDLNMCQIAGMSTLEVRVYHNDLAARNVLYVSPHVKITDGLARLIDIDTEY 875

QY 884 HADGKVPYIKMMALLESILRRFTQSDVMSYGVTVMEIMFGAKPYDGIIPAREIPDLEK 943
DB 876 NADGKMEIKMMALLECIFYHRTQSDVMSYGVTVMEIMFGAKPYDGIIPAREIPDLEK 935

QY 944 GERLPQPICTIDVYIMVYKMMIDSCRPFRFELVSEFSRMARDPQRFVYIQNED-LGP 100
DB 936 GERLPQPICTIDVYIMVYKMMIDADSREFKELAEFSRMARDPQRFVYIQNED-LGP 995

QY 1003 ASPLDSTYRSLDEDDMDGLVDAEVLVPPQGFPCDP----- 104
DB 996 PSPNDKTFQNLDEDEDDMDLBEVLV-QAENIPPLYTSTRIDNSRSEIGHSPP 105

QY 1042 --APGAGVYHRRHRSSTRSGGDLTLGLEPSEEAAPRPLAPSEBAGDVFDDGLMG 105
DB 1055 AVTPMSGQFYVQDGFATQGG--MEMPTATSTIPBAVVA--QCATMEFDDSCNG 110

QY 1100 AAKGIQSLPTHDPELQYSEDPVPLPS-----ETDGVAPLTSPPPEVYNQDVR 115
DB 1110 TLKRVVPHVEDSDTQYSDPVPAPERNPRAELDEBGMTPMDHKPQOEYLNPEV- 116

QY 1153 POPSPPREGPLPAARPAAGATLERAKTLPKNGGVYKDVFAFGAVENPEVLTFGGAPQ 121
DB 1168 -----ENPVSNR-----KNQDLQ-----ALNDPEVHNSASG- 119

QY 1213 PHPPPA-----FSPAFDNLVYWDQDPPEPCA 123

DB 1195 --PRKAEDEVNPELYLNTFTNALGNAYMKNSLLSYPERAKKAFDNDYWNHSLPRSR 1252
 QY 1239 --PSTFKGPT-----AENPEYL 1255
 DB 1253 LQHPDYLOEYSTKTFYKONGRIPIVAENPEYL 1285

RESULT 8
 XMRK_XIPMA STANDARD: PRT: 1167 AA.
 ID XMRK_XIPMA
 AC P1338:
 DT 01-JAN-1990 (rel. 13, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Witbrodt C., Adam D., Maltsechek B., Naeseler W., Raulf F.,
 RA Telling A., Robertson S.M., Scharf M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharf M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL, X16891; CAA34770.2; -
 DR PIR, S06142; S06142.
 DR HSSP: P1362; 1FGK.
 DR InterPro: IPR000494; EGF_R_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; Kinasase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_Ldomain; 2.
 DR PRINTS: PR0109; TRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART; SMO0261; FU; 5.
 DR SMART; SMO0220; S_TKC; 1.
 DR SMART; SMO0219; TYKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 666 1167 POTENTIAL.
 FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 716 724 ATP (BY SIMILARITY).
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 407 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.2%; Score 2688.5; DB 1; Length 1167;
 Best Local Similarity 45.0%; Pred. No. 2e-136;
 Matches 573; Conservative 171; Mismatches 385; Indels 145; Gaps 25

QY 4 AALRGWGLLALLPRGAAS*---QVCTGDMKRLPASFTLDMRLHYOCQVQGN 59
 DB 8 AALLQ--LLVLISIRCCSIDPDRKVCQGSNQWTM--LDNHYLKXKXKMYSCNVVLEN 62
 QY 60 LELTYLPTNASTLFLQDIOEVQGYVLLAHQVQVPLQRLIYRGTLFEDNYALAVLDN 119
 DB 63 LEITYQENQDLSFLQSIQEVGVVLLAHNEVSTIPLVNLRLRQGLVHGNTLLVMSN 122
 QY 120 GDPANNTPTVTSAGSPGIGRELQASLREILKGVLYGRNPQLCYQDTILMKDIFKXNQL 179
 DB 123 YQK-NPSSP--DVYQVGLKQLQSLNLEILISGGKXVSHNPLCNVETIMWMDIVDKTSP 179
 QY 180 ALTLIDYRSRACHPCSPMKGSRCSWSESSESDCOSLRYTCAGGC-ARCGPLPTCCHE 238
 DB 180 TMLILPHAFERQCKCHGCGNCSWAPRGCHQCKFKLLCAQCNRRCGPRPIQCNH 239
 QY 239 QCAAGCTGPKASDGLCLHFNHSGICGLHCPALVYNTDTFFESMPNPEGRTFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFDDTCDTCTPPPKIYDIVSHQVVDNPNIKYTFGAACVXE 299
 QY 299 CPYVYLSIDVSGCTLVLPRLNHOEYTAARDGQOREKSKPCARVCYGIQGMN-NFTVSFWL 357
 DB 300 CPSNIVYTE-GACVRSAGKLEVD-ENGRSCKPCDGVCPKVCDDGIGISLSNTIA--- 354
 QY 358 RVKVSASHLEEFAGCKKIFGSLAFLEPESDGPASNTABLAPEQLQVPELTLEITGYLY 417
 DB 355 ---VNSNINISFNCXTKINDIILNNSPEGDPHYIGTMDEHLMNLTVAKEITGYLY 410
 QY 418 ISANPDSLPDLSTVQNLQVIRGLIHNGAYS-LTLQGLGISWGLSRLSGGLALIH 476


```

Db      411 IMMPEMNTSLSVQNLLEIRGRTTFSGKGFVVVVQVRHLOMGLRLKSVKSNVILKN 470
Qy      477 NTHLCFVHTVMDQLFNNPHQALLHTANRPDECVGGLACHQLCARGHGWGPGPQCNA 536
Db      471 TLQIRYANTIMWRILFRSEDSISDYDART-----ENQTCNNECSDDGCGGPTMCVS 523
Qy      537 CSQFLRQGECEVEECVLOGLPREVYNARHCLPCHPEGCPONGSVTCGPGPADQCVACAHY 596
Db      524 CLHVDGRGRCVASCNULQGEPRERAOVDGRVCVQCHQELVQTDLSITCGPPRANCSAHF 583
Qy      597 KDPFCVARGCPGKPDLSYMPIMKFPDEBAGCCPCPINCTHSCVDLDDKXCPAQRASP 656
Db      584 QDGPQCIRCPCHGILGDDDTL-IMKYADKMGQCCPCPCNQCNGCGSGPGLSGCCRGD-IVSH 641
Qy      657 LLSISAVVGLLVVVGAVGGLIKRQCKIRKYMRLLOEVELEPILPSGAMPNQA 716
Db      642 SSLAVGVSGLLITVALLVLLVLRRLRRRK-RKRITRCLQELVEPLTPSGQAPNQA 700
Qy      717 QMRILKETELRKVYLGSGAFGYKGIWIPDGENVXIPVAIKVLRENTSPKXANKELDE 776
Db      701 FLRIKETEFKKDRVLSGAFGYKGLMNDGENIRIPVAIKVLRATSPPXNQEVLE 760
Qy      777 AYVAGVGSFYVSLGLCTSTVQVLTQMLPAGCCLDHYRENGRGLSGODLIMCQOLA 836
Db      761 AYVAVSVDPHVCRLGLCTLSAVQLVLTQMLPYCCLDYVQHGERLGGWLLMWCVOJA 820
Qy      837 KMSYLEDVRLVHRDLAARNVLVKSPNFKITDFGLARLADIBETEYHAGGKVPKIMWA 896
Db      821 KGMNLEERHLVHRDLAARNVLLKNPNVKITDGLSKLTLADKEQAQAGGVPIKMA 860
Qy      897 LESILRRRFTQSDVWSYGVTVWELMTFGAKPYGICPAEIPDLLEGERLPORPCTTD 956
Db      881 LESILQMTYLTQSDVWSYGVTVWELMTFGSKPYGIDPAKEIASYLENGERLPORPCTIE 940
Qy      957 YVMIMWKCMMIDSECRPFRELVSEFSRMAPDPORFVIONEDLGASPPDSTFYFSLLE 1016
Db      941 YVMITLKCMMIDPSRRPFRELVGEFSGMARDPSRYVLTQ--NLPSLSDRLLFSLLS 997
Qy      1017 DDDMGDLVDAEYVLPQGGFCPPAPAGAGMVHRRSSSTRSGGDLTLGLPSEEEA 1076
Db      998 SDD-DVDADEYVLLPYKRI-----NRQGS----- 1020
Qy      1077 PRSLAPSEAGSDVDFDGLDMGAKGLQSLPTHSPFLQRYSDPPLV-PLPSETDGYVA 1135
Db      1021 -EPCLPTGH-----PVENSITLRNISDPLQNALXKDLQGH-- 1056
Qy      1136 PLTCSPOPEYVNOPDVRPQ-----PSPRE-----GPP-ARAPGATLERAKTLSPG 1182
Db      1057 -----EYVNPQSETSRLSDIYNNVEDLTDGNGPVSLSQEALTNFSRPEYLNTN 1108
Qy      1183 KKGAVKDVAFAGAAENPEYLTPOGGAAPQHPPAPSPAFDNLYNDQPPERGAPPST 1242
Db      1109 QMSL--PLVSSGSDDDPY--QAG-----YQAA-----LPQTALGTGN 1143
Qy      1243 FKGRPTAENPEYL 1256
Db      1144 GMFLDPAENLEYLG 1157

```

```

OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90083234; PubMed=2687875;
RA      Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT      "Isolation and characterization of ERBB3, a third member of the
RT      ERBB/epidermal growth factor receptor family: evidence for
RT      overexpression in a subset of human mammary tumors.";
RL      Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90311312; PubMed=2164210;
RA      Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA      Todaro G.J., Shoyab M.;
RT      "Molecular cloning and expression of an additional epidermal growth
RT      factor receptor-related gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN      [3]
RP      SEQUENCE FROM N.A. (SHORT FORM).
RX      TISSUE=Placenta;
RX      MEDLINE=93282822; PubMed=7685162;
RA      Katoh M., Yazaki Y., Sugimura T., Teraoka M.;
RT      "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT      tyrosine kinase.";
RL      Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC      -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC      (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC      SECRETED (SHORT FORM).
CC      -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC      SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC      -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC      SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS
CC      AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC      PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC      -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M29366; AA35790.1; -
DR      EMBL; M34309; AA35799.1; -
DR      EMBL; S61953; AAB26935.1; -
DR      PIR; A36223; A36223.
DR      HSSP; P11362; IFGK.
DR      Genew; HGNC:3431; ERBB3.
DR      MIM; 150151; -
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_DK_kinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_DK_kinase.
DR      Pfam; PF00069; kinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      ProDom; PD000001; Euk_DK_kinase; 1.
DR      SMART; SMO0261; FU; 3.
DR      SMART; SMO0219; TyKc; 1.
DR      PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR      PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```

Alternative splicing.

FT	SIGNAL	1	19	POTENTIAL.
FT	RECEIN	20	1342	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT	DOMAIN	20	643	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	644	664	POTENTIAL.
FT	DOMAIN	665	1342	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	709	966	PROTEIN KINASE.
FT	NP_BIND	715	723	ATP (BY SIMILARITY).
FT	BINDING	742	742	ATP (BY SIMILARITY).
FT	ACT_SITE	834	834	BY SIMILARITY.
FT	DISULFID	186	194	BY SIMILARITY.
FT	DISULFID	190	202	BY SIMILARITY.
FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	266	BY SIMILARITY.
FT	DISULFID	280	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	552	565	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	141	183	EILSGVIEIKNDLCHMDTIDMRDVIYRDAEIVKDNDR SC -> GQFPMVPSGLTPPOPAQDYLDDDDPRLLTSLASAK VPVTLAAV (IN SHORT ISOFORM).
FT	VARSPPLIC	184	1342	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	560	560	E -> G (IN REF. 2).
FT	CONFLICT	1064	1064	E -> G (IN REF. 2).
FT	SEQUENCE	1342 AA;	148097 MW;	7201E7F66CA374BD CRC64;

Query Match 35.6%; Score 2439.5; DB 1; Length 1342;
Best Local Similarity 40.5%; Pred. No. 4.8e-133;
Matches 533; Conservative 194; Mismatches 455; Indels 135; Gaps 32;

QY	307	DVGSCTLVCPILNQEVAEDGTCRCEKSPCARVCYGLGMFNFTVSFWLRYPKVSASH	364
DB	297	DQTSVCRACFPDKNKEVD-KNGLKXKCEPCGGGLCPKACGCTGSGSRFQ-----VDSSN	34
QY	367	LEEFAGCKKIGSLAPFESFDGPAASNTAPLOPELOQVETLEETIGYLYISAMPDLP	426
DB	348	IDGFNCKRIKIGNDIFITGLNGPFWKIRPALDEKINVFRTYREITGYNISWPMHMM	40
QY	427	DLSEVQNLQVIRGRILNNGAYS-LTGLGLISMLGSLRELSGLALIHNTFLCFVHT	485
DB	408	NFSVYSNLTTIGGSLYNRGSLIMGNLNTVSLGFSLSLEISAGRIYISANQLCYHHS	46
QY	486	VPMQDLFENPQALLHTA-NPEDECYEGSLACHQLCARHCWPGGFTQCVCNSQFLRQ	544
DB	468	LMVTVLALGPTEERLIDIKHNRPRDRCAEGKVCPLCSSGCGWPGGQCLSCNYSRGG	52
QY	545	ECVEECRLQGLPREVYNARCLPCHPEQCPQNSVYCFPEADQCCACAHYDPRCA	604
DB	528	VCVTHCNFLNCEPPEFAHEACFSCHPECPMEGTATQNSGSDTCAQCAHFRDGRHVS	58
QY	605	RCPGSKEDLSYMPIWKPDEGACQCPINCTSHCVLDLDDKGPAPQRA-----SPLTSI	66
DB	588	SCPHGVLG-AKGPITYKPDVQNECRCHENCTGCKGPELODCLGTLVLIGKTHLTWA	64
QY	661	VSAYVGLLVVVLGVFGLIKRQOKIR-KYTMRLLOQTELYEPLTPSGAMPNQAOMR	71
DB	646	LTVLAG--LVVI FMMLGOTFLYWRGRRIQNRAMRYLERGESLEPLDPS-EKANKYLAR	704
QY	720	ILKETELKVVULSGAGFYVYKIMIPDGENYKI PAIVKYLRNTPSPKANKEILDEAVY	77
DB	703	IFKETELKLVLSGVFGYVHKVWLPBEGESTIRPICIKYIEKSRQSGFOAVTHMLA	764
QY	780	MAGVSSPYVSLIGICLTSTVQVLTQMLPYGCLLDHVENRGRISGQDLINMCQIAKGM	831
DB	763	IGSLDHAIVLTLGLCGSSIQVLTQVLPGLSLDHYRQALGPOLLNMGQVIAKGM	824
QY	840	SYLEPVRVYHDLAARNVLVKSPHNYITPGLARLLDIDETENHAGGVPIKMALES	895
DB	823	YLEBHGWNHNLAAARNVLVPSQVAVADGAVDLLPPDDKOLLYSEATPIKMALES	882
QY	900	ILRRFTHQSQDWSYGVTLWELMTFGAKPYDGLPAREIPDLLEKGERLPQPICTIDVM	955
DB	883	IHFQYTHQSQVWSYGVTLWELMTFGAEPYAGRLAIVPDLLEKGERLPQICTIDVM	944
QY	960	IMYKCMWIDSCRRPVELYSESRMAKDDQRFVYVJONEDLGRA---SPLDSTFYNSLE	101
DB	943	VMYKCMWIDENIRPVELANEFTRMARDDPRVLYIKRES-GGGLAPGPPHGLTNKLE	100
QY	1017	DDMDQDLVDAEYLVPOQGFPCPDPAAGAGMYHRRSSSTRSGGDLTLGLEP-SEEE	107
DB	1002	EVELEPELDLDLEAED-----NATTTLSALSLPVGTLNRR	104
QY	1076	APRSPLABEGAGSDVFDGLGMAAKQLSLPTHD-PSPLQRYSEDPVPLP-----S	112
DB	1043	GSQSLSPSSGY-WPMQGNLNGEESCQSAVSSSERCPFRVSLH-----PMRGLIASE	109
QY	1129	ETDGYVA-----PLTSGPOPE-----YVNPQVDRPQPSREBP-----	116
DB	1096	SEEGHVTGSEBELQEKVSKCRSRSRSPRPRGDSAYHSQRHSLLPVTVLSPGLEEED	115
QY	1163	-----LPAARPAQATLERRAKTLP-GNGGV-----KDFAFAGAVENPEVLTPOGAA	121
DB	1156	VNGVYMPDTHLKGTPSSREGTLLSVGLSVLGTBEED-----EVEYNNRRRRRS	120
QY	1211	QPHRPPRPFSAFNLNYWD-----QDPRKGAQPSFTFGTPTAENPEVLT	1255
DB	1208	F-PHPRPSSLEELGYEMDVGSDLSASLSTQSCPLHPVIMPTAGTTPDEDEYEM	1263

RESULT 10
ERB3_RAT
ID_ERB3_RAT
AC_Q62799; Q62955;
STANDARD; PRT; 1339 AA.

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbb-3 precursor (EC 2.7.1.112)
 DE (c-erbB3)
 GN ERBB3
 OS Rattus norvegicus (Rat)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RP [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=8522190;
 RA Heller N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein";
 RL Gene 165:279-284(1995).
 RN (2)
 RP REVISIONS TO 85, 513 AND 565.
 RA Heller N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frommet P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neurogins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC3050.1; -
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recept_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 FT Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 224 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 266
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 SQ SEQUENCE 1359 AA; 147545 MW; 0AA5F2402BDF0F1E CRC64;
 Query Match 34.5%; Score 2365.5; DB 1; Length 1339;
 Best Local Similarity 40.7%; Pred. No. 4,36-119;
 Matches 525; Conservative 174; Mismatches 429; Indels 161; Gaps 34
 QY 3 LAALRMGLLALPFGA---STQVCTGDMKRLPASETHLDMLRHLYOGGVQGN 59
 DB 7 LQVLC---FLSLARGESEMSQAVCEGLTLNGSVGDADNOYQTLKYKECEVWGN 62
 QY 60 LELTFLPTNAGSLFQDIOEVGYVLLAHNQVRQVPLQRLRIVRGTQLFEDNALAVLDN 115
 DB 63 LEIVLTGNADLSFLQWIRVETGVIVANMFSLPLPNLRVVRGTQVYOGKFAIFVM-- 120
 QY 120 GDP LNNTPTVTSAGSPGGLREQLRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHKNNQL 175
 DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGYVIEKNKDKLCHMDTIDMRDITVRV--- 170
 QY 180 ALTLIDTRSRACHPSCPKSGSCWSESSSDCSLRTVCAAGC-ARCKGPIPTDCHE 236
 DB 171 GAETIVKXNGANCPCPCHEVCKG-RCKGPGPDQDILTKTICAPQCNKRCGPPNQCHD 225
 QY 239 QCAAGCTPKKSDCLALHPNHSICELHCPALVTYNTDFFESMPNDEGRYFGASCVTA 298
 DB 230 ECAGGSGSPQDTEFCACRPFNDSACVPCRCPEPLVYKLTFLQLEPNHRTKYGVGVAS 285
 QY 239 CPYNYLTSDVSCITLVGLHNQETALDQGRCKKSPCARVXYGCMNNNTVSPWLR 358
 DB 230 CPNNV-VQGFPCVACPCPDMEVD-KGGLKMCPCGGLCPKACEGSGS-----R 339
 QY 359 VPKVASHLSEFAQCKKIFGSLAFPSFGDPSANTAPQPEGLQVFTLEBTITGLYI 418
 DB 340 YQTVDSNIDPFVNCTKILGNLDLTLGLNVDPNPKIPLADPELVNFRVREITGLYI 399
 QY 419 SAMPSLPSDLISVFNQLQVIRIRIHNAYS-LTQGLGISLGLRLRELGSLALIHNR 477

```

Db      400 QSWPHEHNFVSNLTGGRSLYNKGFSLIMKNLNVLSLFRSLKEISARVYISAN 459
Qy      478 TLGCFVTPMDQLFRNPQALLHTA-NRPEDECEVGLACHOLCARGCMGSGPTQCN 536
Db      460 QQLCYHSHLNTRLRGSEERLDIKYDRPLGEBLAGKVCDDLSSGGCMGSGPQCILS 519
Qy      537 CSQFLRGCEVCEERVLQGLPREVYNAHRCIPEPECOPONGSVTCGFPADQVCAHAY 596
Db      520 CRNRSRGVCVTHCNFLOGSEPREFVHAQCFSCPELPEMGSTGCGSDACACAHF 579
Qy      597 KDPFCVAPRCPGKVPDLSTYPIKWPDEEGACOPDINCTHSC--VDLDKCKRPAEGR 654
Db      580 RDGHCVNNSCHGILG--AKGPIKYDAQNECRPCHEMTCGCGNPELDCGQAEVIM 637
Qy      655 SPLTSIVSAVVGLLVVGVGILIKRQOKIR-KYTRMLLOETELVEPLTPSGAMP 713
Db      638 SKPHLVIAVWG--LAVIMILGSGFLYMRGRIRIONKRAMRYLERGESIEPLDPS-EKA 694
Qy      714 NOAOMRIKETELKRVYLGSGAGCTYKGIWIPDGNVYIPIVAIKYLRNTSPKANKEL 773
Db      695 NKVLARIKETELKRLKVLGSGVGVHKGIMLPEGSIXIPVCIXVIEDKSGRQSGFQAV 754
Qy      774 LDEAVYVAGVSGPYVSRLLGICLTSTVQLVQLMPYGLDHYRNRGRIGSGLINMCM 833
Db      755 TDHMLAVGSLDHAHIVRLGICPGSSQLQVLYOYPLGSLLDHVKQREHITGLPOLLNMGV 814
Qy      834 QIAGMSYLEVRLVHRDLARNLVLSVNHVKITDGLARLLDIDETEVHADGKVPK 893
Db      815 QIAGMYLLEHSHVHDLARNLVLSVNHVKITDGLARLLDIDETEVHADGKVPK 874
Qy      894 WMALESILRRRTQSDVMSYGVTVVMTFCAKPYDGIPIAREIPDLKEGERLPPQPIC 953
Db      875 WMALESIHFGKTYQSDVMSYGVTVVMTFCAKPYDGIPIAREIPDLKEGERLPPQPIC 934
Qy      954 TIDVYVMTVCKWMTDSECRPPREFVSESPKARADPQFVYVIONEDLPASPLDSTFYRS 1013
Db      935 TIDVYVMTVCKWMTDSECRPPREFVSESPKARADPQFVYVIONEDLPASPLDSTFYRS 991
Qy      1014 LLEDDMDGLVDAREVLYPQGFCCPDAPAGAGVWHHRSSSTSGGDLTLGIEPSE 1073
Db      992 VLTREL-----QEALEPEL-----DLDLDEABE 1017
Qy      1074 E-----EAPRSLAPSEG-----AGSDVFDGDLGAKAKGQ 1105
Db      1018 EGLATSLGASLPTGLTRPRGQSLSLSSGVMYNGSLGEACLDASVILGRRQGER 1077
Qy      1106 SLPHDPSPLQRYEDFTVPLPSTDGV---APL-----TC-----SPOPE---Y 1145
Db      1078 PLSLH-PIPRGR-----PASESEGHVTSBAEQEKVSVCRSSRSRSPRPGDSAY 1129
Qy      1146 VNCQDVVRPOPSPRECP-----LPAAPACATLERAKTLSP-GKNGV----- 1187
Db      1130 HSGRHSLLTPVTPISPPGLEEDGNGVYMDTLRKASSREGTSLSVGLSTGIEED 1189
Qy      1188 KDVFAGGAVENPYLTPGGAAPQPHPP 1216
Db      1190 ED-----EVEYVMNRKRGRSP-PRFP 1209

```

```

OC      Muscomorpha; Ephydroidea; Drosophiliidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT      that several genetically defined classes of alleles cluster in
RT      subdomains of the receptor protein.";
RL      Genetics 137:531-550(1994).
RN      [2]
RP      REVISIONS.
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2982499;
RA      Lynne E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RT      hormone binding and kinase domains.";
RL      Cell 40:599-607(1985).
RN      [4]
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC      STRAIN=Oregon-R; TISSUE=Embryo;
RX      MEDLINE=87002474; PubMed=3093080;
RA      Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT      "Alternative 5' exons and tissue-specific expression of the
RT      Drosophila EGF receptor homolog transcripts.";
RL      Cell 46:1091-1101(1986).
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP      ANALYSIS.
RX      MEDLINE=99102120; PubMed=9892502;
RA      Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RT      specification in wild-type, Ellipse, and null mutant Drosophila.";
RL      Dev. Biol. 205:129-144(1999).
RN      [6]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC      STRAIN=Berkeley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
RA      Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA      Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Duthin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Kappen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA      Li X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachle J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong S., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195 (2000).
 RN (7)
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekwanryeong;
 RX MEDLINE=85137938; PubMed=2983332;
 RA Wadsworth S.C., Vincent W.S. III, Blodreau-Wentworth D.;
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor.";
 RL Nature 314:178-180 (1985).
 RN (8)
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "Interallelic complementation among DER/fib alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201 (1991).
 RN (9)
 RP REVIEW
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor.";
 RL Cell 89:13-16 (1997).
 CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMIOGEMOSA
 CC AND VENTRAL NEUROECODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC -I- tyrosine phosphate.
 CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAAS1462.1; -;
 DR EMBL; K03417; AAAS1460.1; -;
 DR EMBL; K03416; AAAS0965.1; -;
 DR EMBL; K03418; AAAS1461.1; -;
 DR EMBL; AF109077; AAD26134.1; -;
 DR EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -;
 DR EMBL; AF109083; AAD26135.1; -;
 DR EMBL; AE003454; AAF46732.1; -;
 DR EMBL; X00293; CA26157.1; -;
 DR EMBL; X78820; CA55523.1; -;
 DR EMBL; X78918; CA55521.1; -;
 DR EMBL; X78919; CA55522.1; -;
 DR PIR; A00640; GQFFE.
 DR HSP; P11362; IFGK.
 DR Flybase; FB00003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TyRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; TYRKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS5001; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein, Receptor; Phosphorylation; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KM Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 FT
 Query Match 29.7%; Score 1965; DB 1; Length 1426;
 Best Local Similarity 32.7%; Pred. No. 1,2e-97;
 Matches 468; Conservative 186; Mismatches 432; Indels 344; Gaps 41
 24 QVCTGTMKRLRLPASPEHLDMLRHLYGCGVQVGNLELYPT-NASLSPLODIOEVQ 82
 100 KICGTSRSLSVSNMHNHNRNLRDYNCTGYDGLKLTWLPNENLDSFLDNIREVIG 159
 83 YVLIANQVQVPLQRLIRYGTQLF-----EDNYLAVALDNGDPLNNTPTVTGASPGGL 137
 160 YILISHVDYKRVKVPFKQIIRGRTLLFSLSVEEKXALFV-----TYSKM 203
 138 RELQLRLYLILKGVVLIQGNPOLCQDTIMIDIRKKNQALALTLIDNRACHPSP 197
 204 YTLFIPRLBVLNQGQVFNHNNYLCHKRTIQWELYSNGTDAYNYVDFAPRECKHE 263
 198 MCKGSRWCGSSSEDCSLFTVCAGCA--RCKGPLTDCHEQCAAGGTGKHSQCLAC 255
 264 SCTHG-CWGEGRKQCFKSLTCSPOCAGRCGPRPRCHLFCAGGTGPTOKDCIAC 322
 256 LHFHSGICELHCPALVTYNTDFEEMPNDEGRYTGASCVTAAPYNYLSTVGSTLVC 315
 323 KNFDEAVSKBCEPKPKVPTTYLTNBEGRVAGATVCECP-GHLLRNGACVRSQ 381
 316 PLHNGEVTAAEDGTORCEKSKPCARVCYGLGFMNFTVAFWLRVPVSAASHLEFAGCKK 375
 382 PDKMDKGE-----CVPCNGCPKPTCPG-----VTVLHAGNIDSPRNTCTV 422
 376 IFGLAFLPSPDGC--DPRASNTA-----PLQPEQLQVETLEITGYLYISAMFDSLDP 427

```

Db 423 IDGNIRIDQTFSGFQDVYANYTWGPRYIPLDPERREVFSTVKEITGYLIEGTHPOFRN 482
Qy 428 LSFQNLQVIRGRILHNGAY-SLTLQGLGISMRLRSIREIGSLALIHNTHLCPHNTY 486
Db 483 LSYFNLETTIGRQLMESMFALAIKVSLSYLSMRNLKQISSSVIYQNRDLCTYSNT 542
Qy 487 PMDQFRNPQALLHTANRPEDECVGEGGLACHQLCARHCWGPGPTQCVNCSQFLRGQEC 546
Db 543 RMPAIOKEPEQKQVWVWENLRADLCENKGTICSDQCNEDGCGAGTQDCLTKQNFENGTC 602
Qy 547 VEECRVLOGLEREYVNAHCLPCHPEQCPONGSVTTCGPEPDQCVAAHAKDPFPCVARC 606
Db 603 IADCGYISNAK--FDNRCKIHPBGR-----TCGAGADHCQBCVHVRDQHCSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PKKNYNDGVCRECHATCGDGTGPRDTIGACTTCNLAIINNDATYKRCCLMDKDPD- 713
Qy 615 SYMPIWK--PDEBACQ-----CP-----NCH----- 638
Db 714 GY--FMEYVHPQEQSLPLAGRAVCRKCHPLCCLCTNYGHEQVCSKTHYKRRECET 771
Qy 639 -----SC-----VDLDKG----- 647
Db 772 ECRADHTDEBQRECFQRHPECNGCTGAGDDCKSCNFKLPDANETFPYVNSTMRCIS 831
Qy 648 -CPAQR-----ASPLTS-----IVSAVGIILVVYLVGVFGI 679
Db 832 KCPLEMRHVNQYTAIGPYCAASPRSSKITANDVMNIFITGAVLVPTICLCV--T 889
Qy 680 LIKROQKIRKYT--MRLLQETELVEPLTPSGAMPQAMRIKETEIAKAVNLGSAF 757
Db 890 YICROKQAKKETVMTALSCGDESEPLRSNIGANLCKRIYKDELKGGVGLGAG 949
Qy 738 GTVYKGINIPDGENVKIPVAIKVLRNTSPKANKELIDEAVYVAGVSPYVSLGICLT 797
Db 950 GRVYGVWVPEGENVKIPVAIKELKSTGABSEEPLEAVYVASEHVALKLAVCS 1009
Qy 798 STVQLVTQLMRYGCLLDHVRNBRGLSGODLLNMCQIAGKSTLEVLVYHDLAARNV 857
Db 1010 SQMMLITQLMPGCLLDYVNRNRKIKSGKALLNSTQIAGKSTLEKRLVHDLAARNV 1069
Qy 858 LVKSPNHYKITDFGLARLLDIDETEVHADGKVPKIMMALESILRRFTQSQVWSYGT 917
Db 1070 LVQTPSLVKIIDFGAKLSSDSNRYKAAGKMKIKMLALECIRNRVFTSKDVMAFGT 1129
Qy 918 VWELMTFGAKPYDGIPAEIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRE 977
Db 1130 IMELTTFQORPHENIPAKDIPDLIEVGIKLEQPEICGLDIYCTLSCWHLDAAMRPTFKQ 1189
Qy 978 LVSEFSMAADPOPFVVIQNEDLG--PASPLDSTFFYSLLDD--DMGDLVAEEYIVP 1032
Db 1190 LTVYAEFAARDPGRKIALPGDKFRLRA-----YISOEKDLIRKLAFTTGSALIAK 1242
Qy 1033 QQGFCDPDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAP-----RSPLAPSEG 1086
Db 1243 PDDYLPQKAPAGPS-----HRTDCT-----DEMFKINRYCXPSPKNS 1281
Qy 1087 AGSDVFDG---DLGMGAKGLQSLPTHDSPLQYSDPTVPLPSETDGVVALTLTSPQ 1143
Db 1282 TGDDERDSSAREVGVGNIR-----LDLPVEDDDVLM--TQCPG 1319
Qy 1144 EYVNPQDVPRQPPSPREGPLPAARPAATLERAKTLSPGKGVGVDFAGGAVENPEYL 1203
Db 1320 NNNNNNN-----NPNQNNMAAVGVAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1204 -----TPQGGAPQHR-----PPAPSP-ARDNLYYMD 1230
Db 1359 LNAQTLGVGESPIPTQITGIPVWGGPGMTMEVKVPMPSSEPTSDSHEYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PR: 634 AA.
AC P00534;
DT 21-JUL-1996 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-D.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor."
RL Cell 41:719-726 (1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, M1006; AAA48763.1; ALT_INIT.
DR PIR, A00643; TVCHLV.
DR PIR, B00643; TVFVLV.
DR HSSP, P11362; IFCK.
DR InterPro, IPR000719; Euk_Dkinase.
DR InterPro, IPR001245; Tyr_Dkinase.
DR Pfam, PF00069; Pkinase.1.
DR PRINTS, PR00109; TYRKINASE.
DR ProDom, PD000001; Euk_Pkinase.1.
DR SMART, SM00219; Tyrc; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.5%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.5e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

Qy 593 CAHYKDPFCYARPSSEVKEDLSMPFWKRPDEBACQPPINICTHSCVDLDDKGPAGQ 652
Db 3 CAHPTDGHCVKACPAVIGENDTL-VMKYADANAVQQLCHPCTGCKGPGLEGCP--- 58
Qy 653 RASPLTIVSAV-V-GILVVVYLVGVFGIILIKRROQKIRKYTMRLLOETELVEPLTPSGA 711
Db 59 NGSKTPSIAGVGVGGLCLVVGIGIGLYARR-HIVKRTLARLLQOEELVEPLTPSGE 117
712 MPNOQRILKTELKRVKVLGSGAGFTYVKGIMPDGENVKIPVAIKVLRNTSPKANK 771

```

```

Db 118 APNQAHRLKTEKTERKVKVLSGSAFGTVYGLWMTPEGEKXKIVALKELREATSPRANK 177
QY 772 EILDEAYVMAGVSGPYVSRLLGICLTSTVQVLVQMLPGCLLDHVENRGRGLSODLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQVLVQMLPGCLLDHVENRGRGLSODLLNW 237
QY 832 CMQIAKMSYLEDEVLRVLDLAARVNLVKSPPNHVXTDFGLARLLDIDETEVHADGSKVP 891
Db 238 CVQIAKGMVYLEERLVRDLAARVNLVKTPOHKITDFGLAKLGDENKYNHAGSKVP 297
QY 892 IKMMALESILRRRFTHSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOP 951
Db 298 IKMMALESILHRIYTHSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPPOP 357
QY 952 ICTIDVYMWKCMWIDSECRPRELSEFSRMAAPQRFVVIQ-NEDGPAASPLDSTF 1010
Db 358 ICTIDVYMWKCMWIDSECRPRELSEFSRMAAPQRFVVIQ-NEDGPAASPLDSTF 417
QY 1011 YRSLEDDMDGLVDAEYLVPOQGFPCPPAPGAGGVHRRHRSSTRGGDLTLGLE 1070
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEFAPRSP-----APSEGAGSVDFDGLMGMAKGLQSLPTHDSPLQRYSEDPYV 1125
Db 450 -----SRTPLLSSLSATSNNSATNCID-----RKQGCHPVREDSFVQRYSSDPTGN 495
QY 1126 LPSET--DGYVAPLTCSPQPEYVNPQVRFQPPSPREGPLFAARPAAGTLERAKTILSPGK 1183
Db 496 FLEESIDOGFL-----PAPEYVNG--LMPKKPS-----TAMVQ 526
QY 1184 NGVAVKDF-----AFGAVENPEVLTPOGGAAPPHPPAPSPAFDNLVY 1228
Db 527 NQIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNOSPAAKTIVFESSPY 578
QY 1229 WDQ-----DPEP-----RGAPSTFGKTPTAENPEYLGDPV 1260
Db 579 WQOSNHQINLDNPDYQODFLPNEIKRGLKLVPAENPEYLRVAAP 625

RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (rel. 01, last sequence update)
DT 21-JUL-1986 (rel. 01, last sequence update)
DE 15-UN-2002 (rel. 41, last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RP MEDLINE=84223957; PubMed=6328658;
RA Debire B., Henry C., Benalissa M., Bisette G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -I- tyrosine phosphate.
CC -I- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.

```

```

CC -I- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; K02006; AAA42394.1; ALT INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06644; TVYOH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SMO0219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R->W (IN REF. 2).
FT CONFLICT 140 140 S->F (IN REF. 2).
FT CONFLICT 146 146 I->V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4.4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16

```

D6		CAHFDGPHCVKACPRAGVLGEENDTL-VWKYADANAIVQDLCHPNCTRGGCKGRGLSGCP----	5
QY		653 RASFLSTYSAAV-GLLVYLVTGVVFGLIKRROOKIRKYTKRRLLDSTELVEPLT79GA	711
D6		59 NGSKTPIAAGVGAVGGLCLTVWVLGI GLYLRSR-HIVKRRTLRLQRELEVEPLT8GE	11
QY		712 MPNOAKRIKETELERKVYVLSGAGATGYKKIMVPDENNYKIPAIIVLRENTSPANK	772
D6		118 APNOAHRIKETEPFKVKYLGFGAFGYKKLMIPREGKTIIPALIELDEAISPAANK	177
QY		772 EILDEAYVMAGVSPYVSRLIGICLTSTVOLVTOLMPYGCLLDHYENRGSLGSODLLNM	831
D6		178 EILDEAYVMASVDNPPVCRLTGICLTSTVOLITQLTMPYGCULDIDYREKHONGSYLLNM	23
QY		832 CMOJAKMSLYEDVRLVHRDILAARNVLYNSPNNYKITDPGLARLLDIETIYHADGKVP	897
D6		238 CVOIAKAMNLEEIRHMVHRDLAARNVLYTPQHVIITDFGLAKQUGADEKEYHAEGKVP	297
QY		892 IKMMALESIIRREFTHQSDDVWSYGYVVELMIFGAKPYDGIAPAEIPDLKEGRLPOPP	955
D6		298 IKMMALESIIRHTYTHQSDVWSYGYVVELMIFGSKPDGIPASISVLEKGRLPOPP	35
QY		952 ICTIDVYMVWKCMWIDSECRPFRELISEFSRMARDPQRFVITO-NEDLCPSAPLDTF	107
D6		358 ICTIDVYMIWKCWMSDADSRPKFRELIASFKMARDPPRYLVIGDERMHLPSPDSKF	417
QY		1011 YRSLLIEDDMGDVLVDAAEYLVQQGFCCDPAPAGAGWVHHRSSYRSCGGGLTGLE	107
D6		418 YRTIMEEDMEDLVDADEYLVPHQGFF-----NSPST-----	444
QY		1071 PSEEARSPUL-----APSEGASDVFDGDLGMAGKGLSFLTHDPSPLQRYSEDPTVP	112
D6		450 -----SRFTLSSLATSNNSATNCIDANG-----H-----	476
QY		1126 LPSTDGYNAPLTCSPERYNQDPVRPPSPRRCGPAPAAPPAGAT-LEBAKTLSPCKN	118
D6		477 -PVREDDFL-----PAPRYVNO-LMPKXSTAMVONQINYVISLTAISXLPIDSRYON	527
QY		1185 GVAWDVAFGCAVENPEYL 1203	
D6		528 -----SHSTAINDPEYL 539	
RESULT 15			
EGRF_CHICK			
ID_EGRF_CHICK	STANDARD;	PRT; 703 AA.	
AC_P1387;			
DT_01-JAN-1990 (Rel. 13, Created)			
DT_01-JAN-1990 (Rel. 13, Last sequence update)			
DT_15-JUN-2002 (Rel. 41, Last annotation update)			
DE_Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)			
DE_Fragment).			
GN_EGRF.			
SOS_Gallus gallus (Chicken).			
OC_Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC_Archosauvia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC_NCB1_TaxID=9031;			
RN_(1)			
RP_SEQUENCE FROM N.A.			
RX_MEDLINE=88261272; PubMed=3260329;			
RA_Lak I., Johnson A., Howk R., Sapp J., Bellot F., Winkler M.,			
RA_Ullrich A., Vennstrom B., Schlessinger J., Gliori D.;			
RA "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,			
RT expression in mouse cells, and differential binding of EGF and			
RT transforming growth factor alpha.";			
RL_Mol. Cell. Biol. 8:1970-1978(1988).			
-FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,			
AND ALSO OF TGF-ALPHA, AMBIRINGULIN, HEPARIN-BINDING EGF, GP30 AND			
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).			
-CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
tyrosine phosphate.			
-SUBCELLULAR LOCATION: TYPE I membrane protein			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-325-339-14

Perfect score: 6852

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeopl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6219	90.8	1259	6 O18735	O18735 canis fami
2	3150	46.0	1209	11 Q9OX70	Q9OX70 ratu
3	3121	45.5	1210	11 Q9EP98	Q9EP98 mus muscu
4	2723	33.7	1165	13 Q9YH40	Q9YH40 xiphophoru
5	2708.5	33.5	1137	13 Q9W656	Q9W656 gallus gal
6	2313	32.8	1328	13 P79754	P79754 fugu rubrip
7	2030.5	29.6	1433	5 Q9B1H9	Q9B1H9 anopheles g
8	1872.5	27.3	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.4	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.1	729	15 Q86712	Q86712 avian rous-
11	1718	25.1	567	15 Q86714	Q86714 avian rous-
12	1697.5	24.8	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1660	24.2	962	15 Q84895	Q84895 avian eryth
14	1645	24.0	545	15 Q85468	Q85468 avian eryth
15	1512.5	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1496.5	21.8	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1275	18.6	1193	5 Q9Y1X8	Q9Y1X8 ephydratia f
18	1200.5	17.5	1368	5 Q23821	Q23821 caenorhabdi
19	1180	17.2	1717	5 Q26566	Q26566 schistosoma
20	1131	16.5	527	13 Q90836	Q90836 gallus gal
21	1007.5	14.7	478	11 Q9ESF0	Q9ESF0 gallus nor
22	947.5	13.8	599	13 Q9PSH2	Q9PSH2 gallus gal
23	906	13.2	165	4 Q14256	Q14256 homo sapien
24	887	12.9	176	11 Q923V5	Q923V5 ratu
25	806.5	11.8	346	13 P11776	P11776 xiphophoru
26	778	11.4	435	5 Q8SZM1	Q8SZM1 drosophila
27	754.5	11.0	311	13 Q99162	Q99162 xiphophoru
28	739.5	10.8	1362	13 Q9PV4	Q9PV4 xenopus la
29	734	10.7	331	4 Q9BUD7	Q9BUD7 homo sapien
30	723	10.6	149	6 Q9BG66	Q9BG66 oryctolagus
31	723	10.6	1671	5 Q9NJV5	Q9NJV5 biomphalari
32	694.5	10.1	1368	13 Q8UW85	Q8UW85 parali
33	689	10.1	1418	13 Q93457	Q93457 scophthalm
34	678.5	9.9	1369	13 Q8UW86	Q8UW86 parali
35	666.5	9.7	1472	5 Q9US58	Q9US58 bombyx mori
36	666	9.7	1412	13 Q8UW84	Q8UW84 parali
37	659	9.6	1358	13 Q73798	Q73798 xenopus la
38	640.5	9.3	1418	13 Q8UW83	Q8UW83 parali
39	636	9.3	1245	13 Q9YGH8	Q9YGH8 scophthalm
40	626.5	9.1	2144	5 Q9VD94	Q9VD94 drosophila
41	621	9.1	1371	11 Q9QV4	Q9QV4 ratu
42	601	8.8	987	11 Q91YH0	Q91YH0 mus musculu
43	598	8.7	987	11 Q99WK2	Q99WK2 mus musculu
44	592	8.6	935	4 Q96L35	Q96L35 homo sapien
45	587.5	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

018735 PRELIMINARY; PRT; 1259 AA.

AC O18735; 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)

DE Erdb-2.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008451; BA02127.1; -.

DR HSSP; P11362; IFGK.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_Dkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_kinase.

DR InterPro; IPR004019; YIP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Kinase; 1.

DR Pfam; PF01030; Recep_L_domain; 2.

DR Pfam; PF02757; YIP; 2.

DR Prodom; P0000001; Euk_kinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TykC; 1.

DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1259 AA; 157989 MW; E37364D49C4ACD46 CRC64;

Query Match 90.8%; Score 6219; DB 6; Length 1259;
 Best Local Similarity 90.7%; Pred. No. 0;
 Matches 1149; Conservative 46; Mismatches 58; Indels 14; Gaps 4;

QY 1 MELALCRWMLLALLPRGASITQVCTGDMKRLPASPETHLDMRLHYGCGVYQGNL 60
 1 MELAACRWMLLALLPRGASITQVCTGDMKRLPASPETHLDMRLHYGCGVYQGNL 60
 DB 1 MELAACRWMLLALLPRGASITQVCTGDMKRLPASPETHLDMRLHYGCGVYQGNL 60
 QY 61 ELTYLPTNASTLFLQDIQEVQGVYVLAHQVROVPLQRLIRYRGTLFEDNVALVLNG 120
 61 ELTYLPTNASTLFLQDIQEVQGVYVLAHQVROVPLQRLIRYRGTLFEDNVALVLNG 120
 DB 61 ELTYLPTNASTLFLQDIQEVQGVYVLAHQVROVPLQRLIRYRGTLFEDNVALVLNG 120
 QY 121 DPLNNTVTGASPGGLRELQLRLSTLTKGQVYLQGNPLCYDTLLMKDIFHKNOLA 180
 121 DPLNNTVTGASPGGLRELQLRLSTLTKGQVYLQGNPLCYDTLLMKDIFHKNOLA 180
 DB 121 DPLNNTVTGASPGGLRELQLRLSTLTKGQVYLQGNPLCYDTLLMKDIFHKNOLA 180
 QY 181 LTLIDTRSRACHPCSPMCKSGRCWGESEDDCSLRTVCAGGACRCGRLPTDCHEOC 240
 181 LTLIDTRSRACHPCSPMCKSGRCWGESEDDCSLRTVCAGGACRCGRLPTDCHEOC 240
 DB 181 LTLIDTRSRACHPCSPMCKSGRCWGESEDDCSLRTVCAGGACRCGRLPTDCHEOC 240
 QY 241 AAGCTGKHSDCIACLHPNHSIGICELHCPALVTYNTDTESPNPEGRYTFGASCVTACP 300
 241 AAGCTGKHSDCIACLHPNHSIGICELHCPALVTYNTDTESPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGKHSDCIACLHPNHSIGICELHCPALVTYNTDTESPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTIVCPLEHNOEVAEDGTQRCCKSCPKARVCYGLGMPNFTVSMRL-V 359
 301 YNYLSTDVGSCTIVCPLEHNOEVAEDGTQRCCKSCPKARVCYGLGMPNFTVSMRL-V 359
 DB 301 YNYLSTDVGSCTIVCPLEHNOEVAEDGTQRCCKSCPKARVCYGLGMPNFTVSMRL-V 359
 QY 360 PKVSASHLEEFACCKIFGSLAFLPESFDGDPASNTAPAPQLOVFTLEBITGYLIS 419
 360 PKVSASHLEEFACCKIFGSLAFLPESFDGDPASNTAPAPQLOVFTLEBITGYLIS 419
 DB 360 PKVSASHLEEFACCKIFGSLAFLPESFDGDPASNTAPAPQLOVFTLEBITGYLIS 419
 QY 420 AMPDLSPLDSVFQNLQYIRRIHNGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 479
 420 AMPDLSPLDSVFQNLQYIRRIHNGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 479
 DB 420 AMPDLSPLDSVFQNLQYIRRIHNGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 479
 QY 440 AMPDSLPLNSVFNLRIRGRVLDGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 473
 440 AMPDSLPLNSVFNLRIRGRVLDGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 473
 DB 440 AMPDSLPLNSVFNLRIRGRVLDGAYSLTLOGLSIWLGLRSIRELQSGIALIHNTH 473
 QY 480 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 539
 480 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 539
 DB 480 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 539
 QY 474 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 532
 474 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 532
 DB 474 LCFVHTYPMQDLFRNPHQALLHTANRPEDCEVGEGLAQCLQARQCHQCPGTQCVNSQ 532
 QY 540 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 599
 540 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 599
 DB 540 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 599
 QY 533 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 592
 533 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 592
 DB 533 FLRGQECVCECRVLOGIPREYVNAHRCPLCPHECOPONGSVTCFPEADQCVACHYKP 592
 QY 600 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 659
 600 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 659
 DB 600 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 659
 QY 593 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 652
 593 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 652
 DB 593 PCVARGSPGVKPDLSYMPIWKEPDEEGACORPINCCHSCVDLDDKCPAQRASPLTS 652
 QY 660 IVSAVGIILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 719
 660 IVSAVGIILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 719
 DB 660 IVSAVGIILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 719
 QY 653 IIAAVGIIILAVVVGIVLIGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 712
 653 IIAAVGIIILAVVVGIVLIGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 712
 DB 653 IIAAVGIIILAVVVGIVLIGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMR 712
 QY 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 779
 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 779
 DB 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 779
 QY 713 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 772
 713 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 772
 DB 713 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKEIIDEYV 772
 QY 780 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 839
 780 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 839
 DB 780 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 839
 QY 773 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 832
 773 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 832
 DB 773 MAGVSPYVSRLLIGICLTSTVQVLVQLMPEYGLLDHVENRGLSODILMCMQIAKM 832
 QY 840 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 899
 840 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 899
 DB 840 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 899
 QY 833 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 892
 833 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 892
 DB 833 SYLEDVRLVHRDLAANVLYKSNHXTIDFGARLLDIDETRYADGGKVPKMALES 892
 QY 900 ILRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 959
 900 ILRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 959
 DB 900 ILRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 959
 QY 893 IIPRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 952
 893 IIPRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 952
 DB 893 IIPRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 952
 QY 960 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1019
 960 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1019
 DB 960 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1019
 QY 953 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1012
 953 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1012
 DB 953 IMYKCMNIDSECPRRRELVSFSRMAPDQSFVVIQNEIDLCPASPLDSTFYRSLIEDD 1012
 QY 1020 MGDLVDAEELVPLQGFPCDPAPAGAGVWHRHSSSTRSGGDLTLGLPSESEARSR 1079
 1020 MGDLVDAEELVPLQGFPCDPAPAGAGVWHRHSSSTRSGGDLTLGLPSESEARSR 1079
 DB 1020 MGDLVDAEELVPLQGFPCDPAPAGAGVWHRHSSSTRSGGDLTLGLPSESEARSR 1079

DB 1013 MGDLVDAEELVPLQGFPCDPAPAGAGVWHRHSSSTRSGGDLTLGLPSESEARSR 1079
 QY 1080 PLAPSGASDVFDGDLGMAKAGIQLPTHPSPLOQRYSEPTVPLPSESDGYAPLTC 11
 1080 PLAPSGASDVFDGDLGMAKAGIQLPTHPSPLOQRYSEPTVPLPSESDGYAPLTC 11
 DB 1073 PLAPSGASDVFDGDLGMAKAGIQLPTHPSPLOQRYSEPTVPLPSESDGYAPLTC 11
 QY 1140 SPQPEVYNQDVARPPSPREGLPLPAARPAGATLER-----AKTLPKNGVYKDVFAFG 11
 1140 SPQPEVYNQDVARPPSPREGLPLPAARPAGATLER-----AKTLPKNGVYKDVFAFG 11
 DB 1133 SPQPEVYNQDVARPPSPREGLPLPAARPAGATLER-----AKTLPKNGVYKDVFAFG 11
 QY 1195 GAVENPEYLTTPQGAAPQPHPPAPSPAFDNLVYWDQDPERGAPSTFKGPTAENPEY 12
 1195 GAVENPEYLTTPQGAAPQPHPPAPSPAFDNLVYWDQDPERGAPSTFKGPTAENPEY 12
 DB 1193 SAVENPEYLTTPQGAAPQPHPPAPSPAFDNLVYWDQDPERGAPSTFKGPTAENPEY 12
 QY 1255 LGLDVPPV 1261
 1255 LGLDVPPV 1261
 DB 1253 LGLDVPPV 1259

RESULT 2
 09QX70
 ID 09QX70 PRELIMINARY; PRT: 1209 AA.
 AC 09QX70;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=90258688; PubMed=2342466;
 RA Petch L.A.; Harris J.; Raymond V.W.; Blaeband A.J.; Lee D.C.;
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 encoded by an alternatively spliced transcript in normal rat tissues."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K.; Dawson T.L.; Earp H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: M37394; AAF14008.1; -.
 DR HSSP: P11362; IFGR.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU_3
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KM ATP-binding; Receptor; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B773 CRC64;

Query Match 46.0%; Score 3150; DB 11; Length 1209;
 Best Local Similarity 50.1%; Pred. No. 52-229;

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 2.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D871A74E CRC64;

Query Match 39.7%; Score 2723; DB 13; Length 1165;
 Best Local Similarity 45.3%; Pred. No. 9, 1e-197;
 Matches 560; Conservative 172; Mismatches 379; Indels 148; Gaps 31;

QY 1 MELAALCRWGLLALALPPG-AAST-----QVCTGDMKRLPASPETHLMDRHHVYGGCV 55
 Db 4 LELLEL-----LILLILSIGRCSTDPDRKVCQGTGNTMT---LDNHLKKKKMYSGCNV 56
 QY 56 VQGNLELTPTPTNASTFLQDIQEVGYLLAHNOVROYPLQRLATVNGTOLFEDNATA 115
 Db 57 VLENLEITTYQENQDLSFQSIQEVGYLLAHNVSTPLVNLRLINGQNLVYGNPFTLL 116
 QY 116 VLNDGDLNNTTPTVGASPGSLRELQSLSLTEILKGYLLIQRPOLCYQDTITLKDI FHK 175
 Db 117 VMSNYOK-NPSS--DYVQGLKQQLSLTETILSGYKVSHPNPLCNVETINMMDLYDK 173
 QY 176 NNQALTLIDNRSRACHPSCPMCKSGRCWGSSESDCSLRTVYVAGGC-ARCKGPLPTD 234
 Db 174 TSNPTMNLIPHAFFERQCKQCDPCGVNCSGMAPGPHCQKFTKLCAEQCNRRCRGPKRID 233
 QY 235 CCECCAGAGCTGPKKIDCLACLHPNHSICELHCPALVTYNTDIFESMPNDEGRTFGAS 294
 Db 234 CNEHCAGAGCTGPRADCLACRDFDDGCTCKTPPKIYIVISHQVVDNINITYTGAA 293
 QY 295 CVTACPYNYLSTDVGSCTLVCPAHQEVYAEADGTORCEKCKSPCARVYGLGMEN-NFTV 353
 Db 294 CYKECPNSVYTE-GACVRSAGAMLEVD-ENKSKSKCPGCGVCPKVDGIGISLNTI 351
 QY 354 SWMLRPKYASASHEEPAGCKKI FGSLSLAFPSFPGDASTNAPLQEPQLOVFTLEIT 413
 Db 352 A-----VNSTINISFSNCTKINDIILNRNPFSGDHYXIGPMDPEHLNMLTTVEIT 404
 QY 414 GYLISAMPDSLPLDISVFONLQVIRGRILHNGAYS-LTLOGLSIWLGRSLRELISGLA 472
 Db 405 GYLIVMMPEMNTSLSVFONLEIIRGRITFSRGSFVYVQVSHQWJLSLKEVSGNV 464
 QY 473 LTHNTHLCFHTVPMDDLFNPHQALHTANRPEDCEVGEGLACHOLCAQHMGWGPPT 532
 Db 465 LKNTPOLRYASTIMWRLLFRSEDSQIEYDART-----ENQCNNECSBDGCMGPPT 517
 QY 533 QCVNCSQFLRGQECVBEGRVQGLPREVYNARHCLPCHPEQPNQSGVTCGEPADQVA 592
 Db 518 MCVSLHYDRGRCVASCNLQGEPRERAOVDGRVQCHQECLELVQTSILTCYGPAPANCSK 577
 QY 593 CAHYDPPFCVAPRCPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 652
 Db 578 CAHPDGFQCI PRCPHGMIGDGTI-IRKYADKMGQCPCHQNTQCGSGSGSLGCGD- 635
 QY 653 RASPLSTISAVVGLLVVYGVNGLIKRQCKIRKTYMKRLLOTELVEPLTPSGAM 712
 Db 636 IVSHSSSLAVSGLLITIVALLIVALLIVLRRLRRRIK-RKRTIRRLLOKEIVEPLTPSGQA 694
 QY 713 PNOQMRLIKETELRKVYLGSGAGFYKGIWIPDGENVXIPAIIVYLRBNTSPKXKE 772
 Db 695 PQOAFRLIKETEFKDKRVLGSGAGTYKGLMNPDPGENIFIPAIIVYLRBNTSPKXKE 754
 QY 773 ILDEAYVMAVGSPPVSLGLCLTSTVQVLTQMLPYGCLLDHAVENRGSLGSDLLMNC 832
 Db 755 VLDEAYVMAVDHPVCRRLGLCLTSAVQVLTQMLPYGCLLDVYRQHERICQGMILMNC 814
 QY 833 MOIAGMSYLEPVRILVHRDLAARNVLYVSPHVAITPTGRLALDIDETVYHDDGKVPPI 892
 Db 815 VOIAGMNYLEPVRILVHRDLAARNVLYVSPHVAITPTGSKLTLADEKTYQAHGKGVPI 874
 QY 893 KMALESLIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRLLPQPI 952
 Db 875 KMALESILQMTYTHQSDVMSYGVTVWELMTFGSKPYDGI PAKELIASVLENGERLPQPI 934

QY 953 CTIDVYMVKCMIDISECRPRFRELVSFSRMAADPQRFVYIQNEDLGPAASPIDSTFYR 10
 Db 935 CTIEVYMILKCMIDIPSSRPRFRELVSFSQMAADPQRFVYIQ--NLPSRDRFLFS 99
 QY 1013 SLLEDDMDKLDVDAEELVLPQGFCCPDPAAGAGMTHRRSSSTRGGGDLTGLLEPS 10
 Db 992 RLSSDD--DVVDADEYLL-----RYKRIN-RQGS-----10
 QY 1073 EEEAPRPLAPSEGAGSDVFDGLGMAKGLSLPTHPDPLQRYSEDPTV-PLPSETD 11
 Db 1019 -----EPCLPENGH-----PVENSIALRYISDPQNALERKLD 10
 QY 1132 GYVALPTSPQPEYVNPQVPEQP-----PSPRE-----GPLP-AARPAGATLERAKT 11
 Db 1053 GH-----EYVNOGSETSSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRREY 11
 QY 1179 LSPKXGVYKDVFAFGAVENREYLTGCGAARQHPAPSPAFDNLYYWDQPREBGA 12
 Db 1103 LNTNONSLL--PLVSSGMDPDY--OAG-----YQALF-----LPQGA 11
 QY 1239 PPSRFKGTPTAENPEYGL 1257
 Db 1138 LTQNGWFLPAENLEYGL 1156
 QY 1239 PPSRFKGTPTAENPEYGL 1257
 Db 1138 LTQNGWFLPAENLEYGL 1156
 RESULT 5
 Q9W6F6 PRELIMINARY; PRT; 1137 AA.
 AC Q9W6F6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERB34.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDBRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.,
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RT embryonic chick hindbrain.";
 RL Mol. Cell. Neurosci. 13:237-258 (1999).
 DR EMBL; AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Euk_Pkinase.
 DR InterPro; IPR001368; TNFR_c6.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recept_L_domain; 1.
 DR Pfam; PF02757; YLP; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW kinase; tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;
 Query Match 39.5%; Score 2708.5; DB 13; Length 1137;

Best Local Similarity 46.8%; Pred. No. 1,1e-195;
Matches 536; Conservative 172; Mismatches 354; Indels 83; Gaps 25;

```

Qy 161 LCYODITLMKDIHKNQALTLITDINBRACHPCSPMKSCRCMGSSEDDCQSLTRTV 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 LCYADITLMKDIHKNQALTLITDINBRACHPCSPMKSCRCMGSSEDDCQSLTRTV 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 221 AGGC-ARCKGPLPTDCEQCAAGCTGPRHSDCLAELHFNHSGICELCPALVTYNTDTF 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 AEQCDRCYGVYSDCHREACAGCGSPKDTDCACMNFNDGACVTCQPPQVYVNPPTF 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 280 ESMRPEGRYTFASCTVACPRVYVSTDVGSCTLVCPHNVQVTVEDTQCEKSKPCA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 QLEHNNAKYTYGAFCKKCPNHFV-VDESSCCVRACPSSKMEV-BENGIKCKCPCTD:CP 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 340 RVCYGLMENNFTVSPMLRVKVSASHLEEFAGCKKIFSGSLAFLEPSEDFPASTAPLQ 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 KACDGI-----TGSIVSAQVYDSSNIDKFNCTKINQNLFLVTGHHGDPYHTIAIN 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 400 PEOLQVETLEITGYLYISAMPDSLDPVSFQNLQVIRGRILHNGAVSLTQGLISWL 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 PEKINIFQVIREITGYLINQSMENMTDERVFSNLTIGGRALVGLSLILKQGLITSL 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 460 GLRSLRELGSGLALHNNHPLCFVHTVPMDOFPRNPHQALHTARPEDECYEGELACHQ 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 QFOSLKOISAGNIYITDNSLYCTHTVNTSLFSTPSQTVVHRKKAENCTADQMCNE 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 520 LCARGHWGPPTQCVNCSQFLRGCEVECRYLOGLPREVYNAHCLPCHBECOP-ONG 578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 LCSSDGCWGPDPQCLSKFRIGRFTIESCNLYDEFEFPANSGVCECDPQCKMEDN 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 579 SVTCFGEADQCAAHYDPFPCVARGSGVYKPRDLSYPIKPFDEGACQPCPINTH 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 MITCYGGPDPCHCKCHFDGPRCAKCDGLQGANSP--IKVYDEBRECHPCHPNCTQ 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 639 SCVDLDKGC-----PAECRASPLTISAVV-GILLVVLGVGEGILIKRQ 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 GCRGPAASHDCTVYPMWRQSTLPQHAR-TFL--IAGVIGGLIYIMGLTFVYARRSI 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 687 KIRKYTRRLQETELVEPLTPSGAMPNOAKRIKETELKRYKVLGSGAGRTYKGIWI 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 K-KKRALRRL-ETELVEPLTPSGAMPNOAKRIKETELKRYKVLGSGAGRTYKGIWI 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 747 PDGENYKIPVAKIVLRBENTSPKANKELDEAYVMAGVSPYVSLGICLSTVQLTQL 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 PBGETYKIPVAKIILNETTGPRKXNVFEMDEBALIMSMDBPHLVRLGICLSTVQLTQL 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 807 MPYGLIDHYRNRGRUGSODLINMCMQIAKGNSTYLEDVLRDLAARNVLVKSPTHK 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 MPHGCLIDVYHEKMDINISQLLIMCVOIAKGMVYLEERLVRDLAARNVLVKSPTHK 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 867 ITDFGLARLIDIDETEHADGGKVPKIMMALESILERRTHQSDVSYGVYVWELMTG 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 ITDFGLARLIDIDETEHADGGKVPKIMMALESILERRTHQSDVSYGVYVWELMTG 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 927 KPYDGIPIREIPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVEFSMA 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 KPYDGIPIREIPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVEFSMA 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 987 RDPORFVYIIONED-LGPASPLDSTFRSLLEDDMDMDLVDAEYLVPOQGFPCRPAGA 1045
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 RDPORFVYIIONED-LGPASPLDSTFRSLLEDDMDMDLVDAEYLVPOQGFPCRPAGA 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1046 GGNVHHRHRSSTRSGGDLTLGLEPSEEARPS--PLAP--SEGASVDFDDLMGAAX 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 RTRIDSVNRQFVYRDGYAAEGV-PMYRABGCIPEAPVAGATAEFEDTCCNGLR 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1103 GLOSLEPHDPSPLQRSSEPTVPLBS-----ETDGIYALPTCSPOPEYVQPVPRQP 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 KOYATLAKEDSSYQRSADPTVFIPIRVIRGELDEGYMTPRDKPDYINPVENFV 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1156 PSPREGPLPAA-PPAGATLEBAKTLSPGKGVKGVF-----AFGAVENPEYLPQ 1206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1005 SRRKNGDLQAVDNPEYH-----APNGQPKADEVNEEYLVNTEANTLENAYL-- 10
Qy 1207 GGAAPQHPPEPAFSAPFNLYWDDDPERGA--PSTFKGTP-----AE 12
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 -----KNMPEKAKQAQFNDPDMHSLPBRSTLGHFDLYQSYSTKYKQNGRIRPIAE 11
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1251 NPEYL 1255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 NPEYL 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE EtdB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 Kb of genomic DNA around the wnt1 locus of Fugu
    rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGR.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM0019; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258647B9 CRC64;

Query Match 33.8%; Score 2313; DB 13; Length 1328;
Best Local Similarity 40.4%; Pred. No. 1,1e-165;
Matches 522; Conservative 155; Mismatches 419; Indels 196; Gaps 3

Qy 9 WGLLALLPP--GAASDQ---VCTGDMKRLPASPETHLMDRLHYOGCQVVGNIETL 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 WRLILCVASRLRAASSQTOEAVCEGTQNGLSSTGSEQNYLVNDRYKGEIINGNIEI 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 63 TYVPTAASIFLQDIOEGVYLIHNVQVPLRLTIVGTQIFENNYLAVLVNDGP 12
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 TQESMWDSFLLTIREVGYLIHNVHFOELPLGQLVIRKNSLYERRPLASVFLN-- 12
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 123 LNNTPVTGASPGRLRELQRLSTELILKGVLLICRNPLQCYODITLMKDIHKNQALAT 18
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ---YKDG--PSGNQGLKMLTEILDOGVQIINNKLRLGPMWYVWDI--RNNDARIE 17
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 183 LIDTNSRACHPCSPMKGSRWGSSEDDCQSLTETVAGGC-ARCKGPLPTDCEHCA 24
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 IQNGERGVCH--KSC-GNYCWGKGKQCQTLITVAPQCNDRCCFTSPRDCCHIECA 22
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 242 AGCTGKHSDCLALHFNHSGICELHCPALVTYNTDTESMRPEGRYTFASCVTACPY 30
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 AGGKGLDIDCFACRLFNDSGACVQCQDTLTKNGTQMETNPPAKYQYISICVSCPT 28
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 302 NYLSTDVSGCTLVCPHNOEYTAEDGTOR-CEKSKSPCARCYGLGNFNNFTVSWFLRVP 360
DB 290 HFV-VDGSSCVSVCPDPMEV--ERGSROCELOCSGCPKCEGTGAEBQRT----- 338
QY 361 KVSASHLEBPAGCKKIFGSLAFLEPSPFGDPAASNTAPLQEPOLQVETLEITGILYISA 420
DB 339 -VDSNINISFINGCTKIQSLHFLVTGILGDPKQVPPADAKKLEVFETVEIDILNIO 337
QY 421 WPDSLPDLVSFQNLQVIRGRILHNGAVSLTQGLGISMGLRSURELGSALAIHNNTHL 480
DB 398 WPKELNDLSVFSLSLTITQGRSLFKRFSLMVMRIFPLTSLGRLSREISDSQVYISQNAHL 457
QY 481 CFVHTVPMQDLFRNH-QALLHTANRPEDCEVGGGLACHQICAGHGWGPGPQCVCNCSQ 539
DB 458 CYHHTVMTQLEFRGSRVANSLSNRPAECVADGRVCDPLCSGSGWGPGPQCCSERN 517
QY 540 FLRGOECVEECRVLOGLPREYVNAH-CLPCHPECPONGSVTCFGEAQCACVAKYD 598
DB 518 YSRHGTQVAGCHFNNGIIPREFAGLVGVACHPECKQOTGASCTGGADECMACIKFRD 577
QY 599 PPFVARGCPGKPLSTMPYWKPPDEGACQCPPICTHSQVLDKCGCPAEPASPLT 658
DB 578 GPYCMSSCPAGVN-DGEXGLIFKEPNREGHCEPHONCTQCSGPGINDC---LEAARLT 633
QY 659 STVASVGLLVVVGAVF-----GILIKRQOKIKRYMRELLQETLEVEPLTPSGA 711
DB 634 ISSGOITGIALCVPAGLIFCLVLPFLGMLHYRGLAIRKRAMRYLESSESFEPLGP-GE 692
QY 712 MPNOQMRILKETELRKVVLGSGAFGTGKGIPIPDGENYKIPVAIKVRENTSPKANK 771
DB 693 KQTKVHARILKPSDRKIKPLGSGVGTGSKFWIPEGETVKIPVALIKTIQDSGRQTF 752
QY 772 EILDEAYMAGVGSYVSRLLIGCLTSTVOLVQTMPCILDLHYRENRGLSQDILNW 831
DB 753 EITDHLMSGSLDHYIVRLIGICGTQLQVTLSSHSLHRIHOKHSLDPQRLNW 812
QY 832 CMQIAKMSYLEDEVLYARDLAARNVILKSPNHKITDFGLARLDDDETEYHADGGKVP 891
DB 813 CVQIAKGVYLEBHRVVKHLLAARNILKNDYQVQIDYVADLTYDDDKVYSEKTP 872
QY 892 IKMMALESILRRRFFHQSVDWSYGVTTWELMTFGAKPYDGIIPAREITDILEKGRLEPP 951
DB 873 IKMMALESILRRRFFHQSVDWSYGVTTWELMTFGAKPYDGIIPAREITDILEKGRLEPP 951
QY 952 ICTIDVYMWYKCMWIDSECRPRELSEPSRMAPORVVIQNDLGPASPLDSTFY 1011
DB 933 ICTIDVYMWYKCMWIDSECRPRELSEPSRMAPORVVIQNDLGPASPLDSTFY 1011
QY 1012 RSLLEDDMDGLVDAEYLVPOQGFCCPDPAPGAGGVHHRSSSTRSGGDLTLGLEP 1071
DB 981 ---EDSGWGEFL-----RGSER---GLEADLEE 1004
QY 1072 SEEEAPRSPLAPSEGASDVFDGLGKW--AAKGLQSLPHDSDPLQ-----RY 1118
DB 1005 DEEB-----GLGDRFATPSLDPSPSMWSTSPQINSYMWMTQLAR 1043
QY 1119 SEDPTVPLPSETDGYVAFLTCSPP-EXVNO-----PVRPQPSPREGL 1163
DB 1044 D-----FAVSGGHHIGYLPMSPSFVDITRQLMWQRSLSVRTLPRSAFRSSRAEL 1097
QY 1164 --PAARPAATLERAKTISPGKNGVMDVFAFGAVENPELITPOGGAALQPHPPAFSP 1221
DB 1098 CEDGAQCAGIFRVR-----FGSERGN-----FOGG----- 1122
QY 1222 AFDNLYWDDODPPERGAPSTFKGPTAENPE 1253
DB 1123 -----QQRKLSTASSPSSFKTMADEDE 1146

```

RESULT 7
Q9BIH9
ID Q9BIH9
AC Q9BIH9; PRELIMINARY; PRT, 1433 AA.

```

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxId=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SDA.
RA Lysate G.U.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSRP; P11362; 1FKK.
DR InterPro; IPR000345; Cytochrome_b.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recp_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Receptor.
KW NON_TER.
FT
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D8896772AF07 CRC64;
QY
Query Match 29.6%; Score 2030.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 2.8e-144;
Matches 470; Conservative 199; Mismatches 393; Indels 391; Gaps 36
QY 26 CTGDTMKRLDASETHLDMLRHYQGCQVYQNLLETTPYTNASLSFLDDIQEVCYVL 85
DB 1 CIGTNGRMSVAPNREHYKMLRDRTYNTCTYVDGDLITWTIONITDINFLQIREVGYVL 60
QY 86 IAHNOVRQVPLQRIYRGTLF-----EDNYALAVLDNDGDPINNTTPVYGASPGGLREL 140
DB 61 ISYLDLPVYLLPRQITIRGRITFKLNMKEEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGSVLIQRNPOLCYODTILMKDI-FHKNOALATLIDNRSRACHPCSPMC 199
DB 105 ELPALRDLIGSVGFNNYVNLCHKVSIINWEILLAPQTSQYTFNFSPRVPCHPCSC 164
QY 200 KGSRCWGSSESSDQSLRTYVAGCA--RCKGPIPDCCHECCAGCTGPKSHDCLALCH 257
DB 165 EVG-CWGSAGANCRFSKLKNSPQCSGRCGPGRPCCHLPAGGCTGTGTQSDCLACKN 223
QY 258 FNHSGICEHLCPALVTYNTDFFESMPNDEGRYTFGASCYACAPYNYLSTDVSGCTLVCP 317
DB 224 FYDDGVCKQCEPPIQVINYPTNYFMEPRPDQKAVAGATCVAKCP-EHLKMGNGACVAKCP 282
QY 318 HNGEYTAEDGTQREKSKPCARCYGLGNFNNFTVSWFLRVKVASHLEEPAGCKKIF 377
DB 283 GKMPQNSE-----CVPKGVCPKTCPEGI-----YHSDNGYKDCDTTIE 323
QY 378 GSLAFLESPFGDPAASNT-----APLQEPOLQVETLEITGILYISAWPDSLPLDS 429
DB 324 GSLRLDQSFQPGQVYTNYSFGFRYKIKIPDRLEVPSTYKELTGFTINIOAHNPFTTLN 383

```



```

QY 430 VFONLOYRGRILHNGAY-SLTQLGLISWLSGRSRELSGLALIHNTHTCFVHTVM 488
DB 384 YFRLNLEVVGGRLKENFASVYIVKXSLKSLKRVNVSIVLENSDLCFVEDIOM 443
QY 489 DQLFNPHQALHTANPEDECEGEGIACHQLCARHGWCPGTQCVNCSQFLRGECE 548
DB 444 SEIKSSDHEVMVQKNNATECHEGECSEOCSSKAGCWAKGEQCLEKNVYKXKCLD 503
QY 549 ECRVLOGI.PREV-VNABHCLPCHPECCOPONGSVTCFSPEDQCVACAHYKDPFCVACP 607
DB 504 SCK--SLPRLYSVDSKTCGDCHQECQ-----FCYPNEDNGSCMANYKDGAFCAECP 555
QY 608 -----SGVXPDLSTYMPWKFPD-- 624
DB 556 TTKAMNGTCINCHKTVCGRGPRDTIAPDGISCDKAIIGSDAKIERCLMKDSCPDGY 615
QY 625 -----EEG----- 627
DB 616 YSDVVLQEEGPLKQISGKAVCRKHPRCKKCTGYGFHEGECQECTGYKKGECDECPD 675
QY 628 -----ACOPCPINCT-----HSCVDL-----DD-----KCGPAEQ-- 652
DB 676 FYANETRICLPCHQECGCHGLDGDHHECNLKLFGDPYDNATPTTCVSNCPASHPYK 735
QY 653 -----RASPLSTYSAVVGLLVVGLGVNGL-----LIKRRQK 687
DB 736 RFPQAGKIGPYCASDMSQSLRIEPTQYKIVGWSVALILCVVGLIAFVLPFRHKNK 795
QY 688 IKKYTMRRLQETELVEPLTPSGAMPNOAKRIKETELRKVYKLSGAGATVYKGIWIP 747
DB 796 KDAVKYTMALAGCESEPLRPSNNGPNTXKRIIKKEIRRGVGLNGAGGRVFKGVWMP 855
QY 748 DGENYKIPVAIKVIRENTSPKANKEILDEAYVAGVSPYSLGLICLSTVOLVQM 807
DB 856 ESESVKIPVAIKVLEMSGSESSKEFLLEAYIMASVEHPNLKTLIAVCMTSQMWMLITQM 915
QY 808 FYGCLLDHYRENGRLGSDLLNMCQIAKGSYLEDVRLVHRDLAARNVLYKSPNHVKI 867
DB 916 PLGCLLDYRNKKKQIKGSKALLNMSTQIRKGMAYIEERLVRDLAARNVLYVCPSPCVKI 975
QY 868 TDFGLARLIDIDETEXHADGKRVPIKMALESILRRFTHQSVDVSYGVVWELMTGAK 927
DB 976 TVFGLAKLIDFDSDEYBAAGCKPIKMLALECIRHVRFTSKDWAFFGITWELTYGAR 1035
QY 928 PYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVYKCMIMISEGRPRRELVEFSMAR 987
DB 1036 PYENVPAKVPPELIEIGHKLPPQDICSDVYICILSCWVLDDARPIYFKQLAETFAEKAR 1095
QY 988 DPQRFVYIQNEDIGPASPLDSTFYRSLLDEDDMDLV----- 1024
DB 1096 DPGRYIMI-----PGDKFMRLPSTYNGDEKOLIRTLAFVMAAAAAAGASAVND 1146
QY 1025 -----DABEYLVPOQGFCCPPAPAGAGMYHRRSSSTSGGDLTLGLEPSEEP 1078
DB 1147 VPSTIAETDEYLOPKRIPRIMLPESA-----VEPS-DEMPK 1182
QY 1079 S-----PLAP--SEGASDVFDGDLGWAAGKQLSLPTHDPSPLOQYSEDPVPLPSE 1129
DB 1183 SLRYCKDPLKPDDETGHGKEV-----GVGGR-----LNLPLD 1216
QY 1130 TDGYVAPLTCSPQPEVYNQPDVPPQPPSPREGPLPAAPACATLERATLSPGKNGVKD 1189
DB 1217 EDDYLMF-TCOSQ-----NOS-----TPG-----YMD 1237
QY 1190 VFAFGAVENPEYL-----TPQGAAPQHPHPAPAFDNLVYWDQPPERGAEP 1240
DB 1238 LIGVPASVDNPEILMSTQAIAGLAGSNG--PHTP-----PF 1274
QY 1241 STEKGTPTAENPE 1253
DB 1275 NTENGMPTHQHSQ 1287

```

```

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9415951; PubMed=10485918;
RA Doherty J.K., Clinton G.M., Adelman J.P., Clinton G.M.;
RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR00494; EGFR_L domain.
DR InterPro; IPR002174; EGF_Like.
DR Pfam; PF00757; Furin_Like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; F6C1BE347E2D030C CRC64;

Query Match 27.3%; Score 1872.5; DB 4; Length 419;
Best Local Similarity 86.5%; Pred. No. 4e-133;
Matches 354; Conservative 6; Mismatches 27; Indels 13; Gaps 1;

QY 1 MELALCEWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYQCCVVOGNL 60
DB 1 MELALCEWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYQCCVVOGNL 60
QY 61 ELTYIPTNASSFLDIOEVGVYLAHNOVROYPLRLRVGTOLPEEDVYALAVDNG 121
DB 61 ELTYIPTNASSFLDIOEVGVYLAHNOVROYPLRLRVGTOLPEEDVYALAVDNG 121
QY 121 DPLNNTPTVGTASPGGLRELOLRSTELIKGVALIORNPOLCYODTILMKDIFPKNNOLA 181
DB 121 DPLNNTPTVGTASPGGLRELOLRSTELIKGVALIORNPOLCYODTILMKDIFPKNNOLA 181
QY 181 LFLIDTNRSPACHPCSPYCKSGRCKGSESSSDCCSLTRVYCGGACGKGLPTDCHEQC 241
DB 181 LFLIDTNRSPACHPCSPYCKSGRCKGSESSSDCCSLTRVYCGGACGKGLPTDCHEQC 241
QY 241 AAGCTGPKKSDCLACLHFNHSGICELHPALVTVYTDFFESMPNPEGRTYFGASCYIACP 301
DB 241 AAGCTGPKKSDCLACLHFNHSGICELHPALVTVYTDFFESMPNPEGRTYFGASCYIACP 301
QY 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVYCYGLGEMNNNTVFWLRVP 361
DB 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVYCYGLGEMNNNTVFWLRVP 361
QY 361 KYASLSLEFFACCKKIFGSLAFLPSPGDPASNAFLQF 400
DB 360 PGPAHFV-----LSFLRPSMDLVSAFYSLPLAP 387

RESULT 9
Q9R2X1 ID Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.

```

OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 4,1e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRFTQSDVSYGVTVWELMTFGAKFYDGIPIAEIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRFTQSDVSYGVTVWELMTFGAKFYDGIPIAEIPDLLEKGERLPQPICT 60
QY 955 IDVYIMVYKCMWIDECRPRRELVSFSSRMARDPQRFVYIQNEDELGPASPLDSTYRSL 1014
DB 61 IDVYIMVYKCMWIDECRPRRELVSFSSRMARDPQRFVYIQNEDELGPASPLDSTYRSL 120
QY 1015 LEDDMGDLVDAEEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDMGDLVDAEEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEE 180
QY 1075 EAPRSPLAPSEAGSDVFDGDLGAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYV 1134
DB 181 EAPRSPLAPSEAGSDVFDGDLGAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYV 240
QY 1135 APLTCSQPPEYVNOVDVPOPPSPREGPLPARPAGATLEBAKTLSPGKGVYKDVFAFG 1194
DB 241 APLTCSQPPEYVNOVDVPOPPSPREGPLPARPAGATLEBAKTLSPGKGVYKDVFAFG 300
QY 1195 GAVENPEYLTQGGAAQPHPPPAFPAFNDLYYWDODPPERGAPSTFTKGTJTAENPEY 1254
DB 301 GAVENPEYLTQGGAAQPHPPPAFPAFNDLYYWDODPPERGAPSTFTKGTJTAENPEY 360
QY 1255 LGLDVFPV 1261
DB 361 LGLDVFPV 367

RESULT 10

Q86712 PRELIMINARY; PRT; 729 AA.

AC Q86712; 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Polypeptide.
GN Polypeptide.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.,
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities."
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -
DR HSSP; P03322; 1A6S
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SMO0219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D256914EF81D63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 3,1e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTFGEADQCAVCAHYKDPFCVACRSGVGFEDSYWIKWFPDEBAGQCPPI 63
DB 141 PEETATPRTGP--DHCKKCAHFIDGPHCVACAPAGVIGENDTL-VWKYADANAVCQLCHP 19
QY 635 NCTHSCVLDLQKGPAGBAGAPLSTISAVY-GILLVAVGVGGLIKRQOKIRKRYTM 69
DB 198 NCTHSCVLDLQKGPAGBAGAPLSTISAVY-GILLVAVGVGGLIKRQOKIRKRYTM 25
QY 694 RRLLETELVEPLTPSGAMPNQAQMRILKETELKRVYVLSGAFYVYKGIWIPDGENV 75
DB 254 RRLLETELVEPLTPSGAMPNQAQMRILKETELKRVYVLSGAFYVYKGIWIPDGENV 31
QY 754 IPIATKULRENTSRKANKEILDEAYVWAGSPVYSLIGICTSTQVLTQMLPYGCL 81
DB 314 IPIATKULRENTSRKANKEILDEAYVWAGSPVYSLIGICTSTQVLTQMLPYGCL 37
QY 814 DHYENRGRSLQDLNWMQIAGKMSYLEDVRLVHRDLAARVLYKSPYHVKITDGLA 87
DB 374 DYIRHKONISQVLLNMCVQIAGKMYLDERRLVHDLAARVLYKTPQHVKITDGLA 43
QY 874 RLDDIDETEHADGGKVPDKMALESLRLRRFTQSDVSYGVTVWELMTFGAKFYDGI 93
DB 434 RLDDIDETEHADGGKVPDKMALESLRLRRFTQSDVSYGVTVWELMTFGAKFYDGI 49
QY 934 AREIPDLLEKGERLPQPICTIDVYIMVYKCMWIDECRPRRELVSFSSRMARDPQRFV 99
DB 494 AREIPDLLEKGERLPQPICTIDVYIMVYKCMWIDECRPRRELVSFSSRMARDPQRFV 55
QY 994 VIQ-NEDELGPASPLDSTYRSLLEDMDGDLVDAEEYLVPOQGFPCDPAPAGAGVYHNR 105
DB 554 VIQ-NEDELGPASPLDSTYRSLLEDMDGDLVDAEEYLVPOQGFPCDPAPAGAGVYHNR 59
QY 1053 HRSSSTRSGGDLTLGLEPSEEAAPRSP--APBEGASDVFDGDLGAKAGLQSL 110
DB 599 NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGQGH 631
QY 1108 PTHDPSLPQYSEDPTVPLPSET--DGTVAPLTCSQPPEYVNOVDVPOPPSPREGPLPA 116
DB 632 PTHDPSLPQYSEDPTVPLPSET--DGTVAPLTCSQPPEYVNOVDVPOPPSPREGPLPA 675
QY 1166 ARPAGATLERAKTLSPGKGVYKDV-----AFGAVENPEYL 1203
DB 676 -----TAMVQNCIYNNISLTAISKLPMDRQYQNSHSTAVDNEYL 715

RESULT 11

Q86714 PRELIMINARY; PRT; 567 AA.

AC Q86714; 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-erbB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,

RA Johnson A, Beng H.
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60727.1; .
 DR HSSP: P1362; PFGR.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase; 1.
 DR PRINTS: PRO0109; TYRKINSE.
 DR ProDom: PDO00001; Euk_kinase; 1.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Tyrosine-protein kinase.
 FT NON_TER
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;
 Query Match 25.1%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 3e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;
 QY 584 GPEADQCAACAHYKPPPCVACRPSGVKPDLSYMPWKFPDEBACQPCPINCTHSCVDL 643
 DB 1 GP--DHCKCAHFIDGPHCVACRPAAGVIGENDTL-VMKYADANAVCCLCHPNCRTGCKGP 57
 QY 644 DEKGPACGRAPLRSIVSAVY-GILVWVGVVGGIIRKQCKIRKYMRLLOETEL 702
 DB 58 GLEGP---NSKTSIAAGVGGILCLVVGGLGILYLR -HIVKRLRLLOREL 113
 QY 703 VEPITPSGAMPNQAQRIKTELKRVYLGSGAFVYKGIWIPDENYKIPVALKVL 762
 DB 114 VEPITPSGAPNQAALRIKTEFEKRVYLGSGAFVYKGLWIPDEKXIPVALKELR 173
 QY 763 EHTSPKANKELDEYVNAVGSPPVSRLLGICLTSTQVLTQIMPGCLLDHRENGR 822
 DB 174 EATSPKANKELIDEYVNAVSDNPRVCLGLCLSTVQLTLQIMPGCLLDYRHKDN 233
 QY 823 LGSQDLNMCQIAKMSYLEVRLVHRDLAARNLVKSPNHVXITDFGLARLIDIDE 882
 DB 234 IGSQYILNMCVQIAKMSYLEVRLVHRDLAARNLVXTPQHVXITDFGLAKLGADEKE 293
 QY 883 YHADGKVPKMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLE 942
 DB 294 YHAEKGKVPKMALESILRRFTHQSDVWSYGVTVWELMTFGSKPYDGIPIAREISVLE 353
 QY 943 KGERLPQPICTIDYVIMVCMIDSECPREFRELVESEFSMARDDPQRFVVIQ-NEDLG 1001
 DB 354 KGERLPQPICTIDYVIMVCMIDSECPREFRELVESEFSMARDDPQRFVVIQ-NEDLG 413
 QY 1002 PASPLDSTFYSLLDEDDMGDIYDAEYLVPOQGFPCDPAPGAGMVAHHRSSSTRSG 1061
 DB 414 LPSPTDSKFRYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
 QY 1062 GGDLTGLPESEEAAPRPL-----APSEAGSDVDFDGLGMAKGLQSLPTDPSPLQ 1116
 DB 455 -----SRTPLLSLATSNNKATNID-----RNGQGHVAVEDSFVQ 491
 QY 1117 RYSEDTFVLPSET--DGIVAPLTCSPQPEYVNDVRRPQPSREGPLPAAPAGATLE 1174
 DB 492 RYSDPTGNFLEESIDDFL-----PAPEYVND--LMPKKPS----- 526
 QY 1175 RAKTLSPGKGVVXDF-----AFGAVENPEYL 1203
 DB 527 ----TAMVONQIYNNISLTAISKLPMSDRYQNSHTAVDNEEYL 566
 RESULT 12
 Q8WYVO
 ID Q8WYVO; PRELIMINARY; PRT; 412 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PP3559.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Wu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318349; AA55956.1; .
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP motif.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PDO00001; Euk_kinase; 1.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 6.6e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4
 QY 895 MALBSILRRFTHQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPICT 954
 DB 1 MALBSILRRFTHQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPICT 60
 QY 955 IDVIMVCMIDSECPREFRELVESEFSMARDDPQRFVVIQ-NEDLGPAAPLSTFVRL 101
 DB 61 IDVIMVCMIDSECPREFRELVESEFSMARDDPQRFVVIQ-NEDLGPAAPLSTFVRL 120
 QY 1015 LEDDDMGDIYDAEYLVPOQGFPCDPAPGAGMVAHHRSSSTRSGGDTGLPESEE 107
 DB 121 LEDDDMGDIYDAEYLVPOQGFPCDPAPGAGMVAHHRSSSTRSGGDTGLPESEE 160
 QY 1075 EAPRSPPLAPSEAGSDVFDGLGMAKGLQSLPTDPSPLQRYSEDPFVLPSETDGYV 113
 DB 181 EAPRSPPLAPSEAGSDVFDGLGMAKGLQSLPTDPSPLQRYSEDPFVLPSETDGYV 240
 QY 1135 APLTQSPQPEYVNDVRRPQPSREGPLPAAPAGATLEPAKTLSPGKGVVXDFVAFG 115
 DB 241 APLTQSPQPEYVNDVRRPQPSREGPLPAAPAGATLEPAKTLSPGKGVVXDFVAFG 300
 QY 1195 GAVENPEYLTQGAAPQ-----HPPA---PSPADNL 122
 DB 301 GAVENPEYLTQGAALSPILLPSAQSTTSINTGRTHQSGGLHAPSKGHLRQRTQST 360
 QY 1227 YVWD-QDPPER-----GAPPSFKTPTPAEN 1251
 DB 361 VWTQCBCEQGVRRSPDVSSGREGLTSAQIKRKEGPPPTTSRGTCHARN 410
 RESULT 13
 Q64895
 ID Q64895; PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gag-V-erb-A-v-erb-B protein.
 GN GAG-V-ERB-A-V-ERB-B
 OS Avian erythroblastosis virus.

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Brustkin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: X52211; CAA36459.1; JOINED.
DR HSSP: P10828; 2NLL.
DR InterPro: IPR00719; Euk_pkinase.
DR InterPro: IPR000536; Hormone_rec_11g.
DR InterPro: IPR001723; Steroid_receptor.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00105; zt-C4; 1.
DR PRINTS: PR00398; STRDHOMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000001; Euk_pkinase; 1.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOL1_1.
DR SMART: SM00219; Tyrc; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;
Query Match 24.2%; Score 1660; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1.6e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;
QY 547 VEECRVLQGLPRE-VYNAR-HCLP-----CHPECC 574
DB 354 IEKCESEYLLAFEHYINRKNIHFWSKILMKVADLMTGAYHASFILMKVECTELS 413
QY 575 PONGSVTCFGEADQCVACAHYKDPFCVACPSGVKPDLSYPIKMFPEEGACQCPPI 634
DB 414 PQE-----VGP--DHCKKCAHFIDGPHCVKACPGVIGENDTL-VWKYADANAQCQLCHP 465
QY 635 NCTHSCVDLDDKGGPAEORASPLTSIVSAVY-GILLVVVLGVVFGILIKRQOKIRKYM 693
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAGVVGGLCLVVVGLGILYLRR-HIVKRTL 521
QY 694 RLLQETELVEPLPLPSGAMPNQAOMRLIKETELAKYVVLSSGAFGVYKGIWIDGENYK 753
DB 522 RLLQERELVEPLPLPSGAPNQAHLRIKETEFKRYKVLGFGAFGVYKGLWIPGEKXT 581
QY 754 IPVAIKVLRNTSPKANKELIDEAAYVAGVSPVVSRLIGICTSTVQVLTOLMPYGCIL 813
DB 582 IPVAIKELRENTSPKANKELIDEAAYVWASVDNPHVCRLLGICLTSTVQVLTOLMPYGCIL 641
QY 814 DHVENGRLGSLQGLLWCMQIAKNGSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLA 873
DB 642 DYIREHKDNIQSQVLLWMCVQIAKGNVYLEERHMYHDLAARNVLYKTPQHYKITDFGLA 701
QY 874 RLDDIDETEHADGKVPKIMMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDGIIP 933
DB 702 KQLGADKEHYHAEQKPIKMALESILHRIYTHSDVMSYGVTVWELMTFGSKPYDGIIP 761
QY 934 AREIPDLLEKGERLPPPICTIDYIMIVKCMWIDSECRREFELVSEPSRYARDPQREYV 993
DB 762 ASEISVLEKGERLPPPICTIDYIMIVKCMWISGADSPKREFELVSEPSRYARDPQREYV 821

QY 994 VIO-NEDLPASPLDSTFYRSLIEDDDMGDLVDAEYLVPOQGFCCPPAPAGAGMTHR 10
DB 822 VIQGERMHLPSPTSKRYRTIMEEDMEDVDADEYLVPOQGF----- 86
QY 1053 HRSSTREGGGLTLGLEPSEEARSPLABSEGSDVDFDGLMGAKQLGSLPTHP 11
DB 867 -NSPST-----SRTPLLSLATSR-----NSATKCIDRNGCH-- 89
QY 1113 SPLQVSDPTVPLPSFTDGVAPLTCSPQEVYNOQPVRRQPPSPREGPLPARPACAT 11
DB 899 -----FREDGFL-----PAEEVYNQ--LMPKKSSTAMVQNOIVYISLT 93
QY 1173 -LERAKTLSPGNKGVKDVFAFGAVENPEYL 1203
DB 937 AISKLPMDSRVQN-----SHSTAVNDPEYL 961
RESULT 14
Q85468
ID Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Avian erythroblastosis virus (T834) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scottling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-276(1987).
DR EMBL: X06943; CA830024.1; -.
DR HSSP: P11362; TRCK.
DR InterPro: IPR00719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCE8C0A0F8AF4 CRC64;
Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 9.3e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15
QY 584 GREADQCVACAHYKDPFCVACPSGVKPDLSYPIKMFPEEGACQCPINTHSCVDL 643
DB 1 GP--DHCKKCAHFIDGPHCVKACPGVIGENDTL-VWKYADANAQCQLCHPNTTRCKGP 57
QY 644 DDKCPAEORASPLTSIVSAVY-GILLVVVLGVVFGILIKRQOKIRKYMRLQETEL 702
DB 58 GLEGC---NGSKTPSIAGVVGGLCLVVVGLGILYLRR-HIVKRTLRLQEREL 113
QY 703 VEPPLPSGAMPNQAOMRLIKETELAKYVVLSSGAFGVYKGIWIDGENYKIPVAIKVLR 762
DB 114 VEPPLPSGAMPNQAOMRLIKETELAKYVVLSSGAFGVYKGIWIDGENYKIPVAIKELR 173
QY 763 ENTSPPKANKELIDEAAYVAGVSPVVSRLIGICTSTVQVLTOLMPYGCILDHVENRGR 822
DB 174 EATSPKANKELIDEAAYVWASVDNPHVCRLLGICLTSTVQVLTOLMPYGCILDYIREHKON 233
QY 823 LGSODLILWCMQIAKNGSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLALDDIDE 882
DB 234 IGSQVLLWMCVQIAKGNVYLEERHMYHDLAARNVLYKTPQHYKITDFGLAQGLADEKE 293

```

QY 883 YHADGGKVPKMMALSLRRRTTOSDVMSYGVTWELMTGAKYDGPAPREIDPLE 942
DB 294 YHAEKGKVPKMMALSLRRRTTOSDVMSYGVTWELMTGAKYDGPAPREIDPLE 353
QY 943 KGERLPQPICTIDVYMWKCMMDSECRPREPRLVSEFSMSAADPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPICTIDVYMWKCMMDSECRPREPRLVSEFSMSAADPQRFVVIQ-NEDLG 413
QY 1002 PASPLDSTFYSRLLEDDMDGDVADAEYVVPQGFCEPDPAAGAGVHHHRSSTTSG 1061
DB 414 LPSPTDSKFRITLMEEDMEDIVDADEYVLPQGF-----NSPT--- 454
QY 1062 GGDLLTGLBPESEABRSP-----APSEAGSDVDFDGLGMAKQGLSPTHDPSFLQ 1116
DB 455 -----SRTPLLSLATSNKSNATNCIDRNG-----H----- 481
QY 1117 RYSEDPVPLPSTIDGVAPLTCSPQPEVYNQPDVPRQPSRREGPLAARPAQT-LEK 1175
DB 482 -----PVREDFPL-----PAPEYVNO--LMPKKPSTAMVQIQIYVLSITLAIK 523
QY 1176 AKTLSPKNGVNDVPAFGAVENPEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNPPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (Tremblre). 12, Created)
DT 01-NOV-1999 (Tremblre). 12, Last sequence update)
DT 01-JUN-2002 (Tremblre). 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RA STRAIN=C97BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheil C.,
RA Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibie N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampard A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA STRAIN=C97BL/6J; TISSUE=LIVER;
RA MEDLINE=21085660; PubMed=1217851;
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batcatov S., Cavaent T.,
RA Fetschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schraml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

```

PA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
PA Gustinich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
PA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombereis P.,
PA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
PA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.F.,
PA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilting L.,
PA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
PA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AF124513; AAD44149.1; -.
DR EMBL: AF275366; AAG28047.1; -.
DR EMBL: AF275364; AAG28047.1; JOINED.
DR EMBL: AF275365; AAG28047.1; JOINED.
DR EMBL: AK004944; BAB23688.1; -.
DR EMBL: AK004883; BAB23641.1; -.
DR EMBL: AK004911; BAB23662.1; -.
DR MGI: 95294; Egfr.
DR WGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FGF_3.
KM Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.1%; Score 1512.5; DB 11; Length 655;
Best Local Similarity 44.2%; Pred. No. 1.3e-105;
Matches 286; Conservative 101; Mismatches 227; Indels 33; Gaps 14

QY 11 LIALIPGAA--STQVCTGDMKRLPASBTHLDMRLHYYQGVVQNLBELTYPTN 68
DB 14 LITLCAAGALAEKKVCGTNSRLTQGTEDHFLSLQRMYNNECEVLGLLEITYQVRN 73
QY 69 ASLSFLQD:QEVGYVLLAHNOVROVPRQRIYRGQLQFEDYVALAVLDGDLNNTTP 121
DB 74 YDLSFKITQEVAGVYLLANTVEIPLENQIIRGNALYNTALALISL----- 12
QY 129 VTGASPGGLREQLRLSTLEILKGYLLIGRNPOLCYODITLWKDI----FKRNNALTLI 18
DB 125 -YGNRTGLRELPMNLOEILLIGAVFSNNPILCNMDITIQRDIVQVFNMSMDL--- 18
QY 185 DNRSGACHPGCPMKGRCWSESEDDQSILTRYCAGGA-RCKGLPTTCHEGCAAG 24
DB 181 -QSHPSCKCPSPGNSCWGGSENGQKLTKITCAQCSHRGRGSPSCCHNCAAG 23
QY 244 CTGPKHSDCLACLNHNSGICELCPALVTYNTDTPESMPPEGRYTFGASCYACPEYNY 30
DB 240 CTGPRESDCLVCQKQDQATCKDTPPLMLVNPPTYQMDVNPBKYSFGATCVKCRNY 29
QY 304 LSTDVGSCTLVPLRNOEYTAEDGTQRCCEKSKPCARVGYGLM--FNNFTVSFWLAVPK 36
DB 300 VYTDHSCVCRAGGPDYEV-EEGDIRKCKKCDGPRKVCNIGIGGEFD-TLS----- 35
QY 362 VASHLSEFPAGCKITFGSLAFLPESFDDPSANTAPLOPELOVETLEETGLYISAW 42
DB 351 INATNIKHYKTALISGLHLIPAFKDSSTRPRDLPRLEILKTYKETGTLGLQAM 41
QY 422 PSLPDLVFNQLQVIRIRILHNGAVSLTQGLISWLGRLSRLSELGSLALHHNTLCL 48
DB 411 PNMVTDLAFENLLEIRRTXQHGQFSLAVGLNITSIGRLSKIEISDGVYIIGNNLCL 47
QY 482 FHTVPMQDLFNPQIALHTANPREDCVGEGLAQGLCARGHCWMPGPTQCNVCSQFL 54
DB 471 YANTTNMKKLTFTPOQTKIKMNRAEKDCYAVNHVNCPLGSSBSCWGPBEPDVCSCNVS 53
QY 542 RQGECEVEGRVVLQGLPREYVNAHRLCPCHPECPQNGSVTCFGEADQCAVAYKDPF 60
DB 531 RRECEVEKKNILGEPBPFVENSEICQHPBCLPQAMNITCTGKGPNNICQCAHYIDGPH 59
QY 602 CVARPGGVKPLDSMPYMKRPDEGACQRPINTCTSSCYDLDDKGC 648
DB 591 CVKTPAGIMGEINNTL-VWKYADANNVCHLCHANCTYCGAGPGLQGC 636

```

Tue Jul 22 12:40:30 2003

seq4-325-339-14.rpt

Page 1

Search completed: July 22, 2003, 09:01:06
Job time : 53.5887 secs

FT /label= Cysteine_rich_domain
 FT Region 210..224
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 250..264
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Domain 324..483
 FT /label= ligand_binding_domain
 FT Region 325..339
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 359..383
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 455..479
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Domain 484..623
 FT /label= Cysteine_rich_domain
 FT Region 579..593
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Domain 624..654
 FT /label= Transmembrane_domain
 FT Region 632..652
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 653..667
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Domain 655..1010
 FT /label= Tyrosine_kinase_domain
 FT Region 661..675
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 695..709
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Region 710..730
 FT /label= insertion region
 FT /label= suitable for foreign epitope insertion"
 FT Domain 1011..1123
 FT /label= C-terminal_domain
 PN WO200020027-A2.
 PD 13-APR-2000.
 PF 05-OCT-1999; 99WO-DK00525.
 PR 05-OCT-1998; 98DX-0001261.
 PR 20-OCT-1998; 98US-0105011.
 PA (MEBI-) M & E BIOTECH AS.
 PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX MPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-196; 220pp; English.
 CC This is the human heregulin 2 (Hetz) sequence. Immunogenic analogues of
 CC Hetz can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Hetz) and/or fibroblast growth factor 8b (FGFB).
 CC The method comprises effecting simultaneous presentation of antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Hetz and human/marine FGFB comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGFB and Hetz, respectively.
 XX

SO Sequence 1255 AA;

Query Match 98.3%; Score 6736; DB 21; Length 1255;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 2;

QY 1 MELALACRMGLIALTPPGASTOVCTGDMKRLPASPETHLMLRHLYOGCQVVGNL 60
 DB 1 MELALACRMGLIALTPPGASTOVCTGDMKRLPASPETHLMLRHLYOGCQVVGNL 60
 QY 61 ELTYLPTNASISFLDDIOEVGYVLIANQVROPLORLRVSTQGFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASISFLDDIOEVGYVLIANQVROPLORLRVSTQGFEDNYALAVLDNG 120
 QY 121 DPLNNTPTPTGASPGGLRELOLRSTELTKGVLIQNNPOLCYDDTLKMDIFHKNNOLA 180
 DB 121 DPLNNTPTPTGASPGGLRELOLRSTELTKGVLIQNNPOLCYDDTLKMDIFHKNNOLA 180
 QY 122 DPLNNTPTPTGASPGGLRELOLRSTELTKGVLIQNNPOLCYDDTLKMDIFHKNNOLA 180
 DB 122 DPLNNTPTPTGASPGGLRELOLRSTELTKGVLIQNNPOLCYDDTLKMDIFHKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPCKGSRGCESEEDCQSLTFTVCAGCARCKGRLPTCCHEQC 240
 DB 181 LTLIDTNSRACHPCSPCKGSRGCESEEDCQSLTFTVCAGCARCKGRLPTCCHEQC 240
 QY 241 AAGCTGPHSDCLACLHNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPHSDCLACLHNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YVYLTDPVSGCTLVCPLNQEVTAEDDQREKSKCARCYGLGMFNPTVSFWLR-V 359
 DB 301 YVYLTDPVSGCTLVCPLNQEVTAEDDQREKSKCARCYGLGMFNPTVSFWLR-V 359
 QY 301 YVYLTDPVSGCTLVCPLNQEVTAEDDQREKSKCARCYGLGMFNPTVSFWLR-V 359
 DB 301 YVYLTDPVSGCTLVCPLNQEVTAEDDQREKSKCARCYGLGMFNPTVSFWLR-V 359
 QY 360 PRYSASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEITGYLYIS 419
 DB 360 PRYSASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEITGYLYIS 419
 QY 354 RAVTSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEITGYLYIS 413
 DB 354 RAVTSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEITGYLYIS 413
 QY 420 AMPDLSLPDLSVFNQVIRGRILNNGAYSLTLOGLSWLGRSLRELGGALIHNTH 479
 DB 420 AMPDLSLPDLSVFNQVIRGRILNNGAYSLTLOGLSWLGRSLRELGGALIHNTH 479
 QY 414 AMPDLSLPDLSVFNQVIRGRILNNGAYSLTLOGLSWLGRSLRELGGALIHNTH 473
 DB 414 AMPDLSLPDLSVFNQVIRGRILNNGAYSLTLOGLSWLGRSLRELGGALIHNTH 473
 QY 480 LCFVATVPMDCLEFRNPHQALHTANRPEDECVGSLACHQLCARGHGMGPGTCVNCQ 539
 DB 474 LCFVATVPMDCLEFRNPHQALHTANRPEDECVGSLACHQLCARGHGMGPGTCVNCQ 533
 QY 540 FRRGQCEVCEGCAVLOGLPREYVYNAHCLCPHEPCQPONGSYTTCGPADQCAAHYKDP 559
 DB 534 FRRGQCEVCEGCAVLOGLPREYVYNAHCLCPHEPCQPONGSYTTCGPADQCAAHYKDP 553
 QY 600 PFCVAPRCPGSGVCPDLSYWPRIKPFDEBGAOCPGICNTHSCVDLDDGCGABRASPPTS 659
 DB 594 PFCVAPRCPGSGVCPDLSYWPRIKPFDEBGAOCPGICNTHSCVDLDDGCGABRASPPTS 653
 QY 660 IYSAVVGLLVVIGVFGILIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNOQMR 719
 DB 654 IYSAVVGLLVVIGVFGILIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNOQMR 713
 QY 720 ILKETELKRVVLVSGAGTGYKGIWIPDGNVYKIPAIIVLSENTSPKXKXILLDEAYV 779
 DB 714 ILKETELKRVVLVSGAGTGYKGIWIPDGNVYKIPAIIVLSENTSPKXKXILLDEAYV 773


```

QY 780 MAGVSPYVSRLLGICLTSTVQLVLTQMLPYGCLLDHVENKRGLSQDLLNMCQIAKGM 839
DB 774 MAGVSPYVSRLLGICLTSTVQLVLTQMLPYGCLLDHVENKRGLSQDLLNMCQIAKGM 833
QY 840 SYLEDVRLVHRDLARNLVKS PNHVKITD FGLARLLDIDETEHADGKVPITKMALES 839
DB 834 SYLEDVRLVHRDLARNLVKS PNHVKITD FGLARLLDIDETEHADGKVPITKMALES 833
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEXGERLPQPFICTIDVYM 959
DB 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEXGERLPQPFICTIDVYM 953
QY 960 IMVXCMIMDSECRPRFRELVSFBSMADDPQRFVYIQNEDIGPASPLDSTFYRSLLEDD 1019
DB 954 IMVXCMIMDSECRPRFRELVSFBSMADDPQRFVYIQNEDIGPASPLDSTFYRSLLEDD 1013
QY 1020 MGDLYDAEELVLPQOGFCPPDPAPGAGMVRHRSSTRSGGDLTLGLPESSEAPRS 1079
DB 1014 MGDLYDAEELVLPQOGFCPPDPAPGAGMVRHRSSTRSGGDLTLGLPESSEAPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAAKGIQSLPTHDPSPLOKRYSEDPTVPLPSETDGVAPLTC 1139
DB 1074 PLAPSEGAGSDVFDGDLGMAAKGIQSLPTHDPSPLOKRYSEDPTVPLPSETDGVAPLTC 1133
QY 1140 SPOPEYVNOPDVRPQPSRREGPLPAARPAATLEBAKTLS PGKNGVYKDVFAFGAVEN 1199
DB 1134 SPOPEYVNOPDVRPQPSRREGPLPAARPAATLEBAKTLS PGKNGVYKDVFAFGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAAPPSTFKGPTAENPEYIGLDV 1259
DB 1194 PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAAPPSTFKGPTAENPEYIGLDV 1253
QY 1260 PY 1261
DB 1254 PY 1255

```

RESULT 2
AAE12130
ID AAE12130 standard; Protein, 1255 AA.

```

XX AAE12130;
AC 18-DEC-2001 (first entry)
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
XX Therapeutic compound; major histocompatibility complex; vaccine;
XX antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
XX adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
XX antigen presenting cell; human; tyrosine kinase-type receptor.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 774..782
XX /note="Antigenic epitope"
XX
XX W020016677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX
XX N-PSDB; AAD19731.
XX

```

PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT

Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).

XX Sequence 1255 AA;

Query Match 98.3%; Score 6736; DB 22; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 2;

```

QY 1 METALCRWGLILLALPPGASTVCGTGMKRLPASPETHLDMRLHLYOGQVVGNTL 60
DB 1 METALCRWGLILLALPPGASTVCGTGMKRLPASPETHLDMRLHLYOGQVVGNTL 60
QY 61 ELLYPTNASTLFLDIOEVGYVLIANQVRQVPLQRLIVRGTOLEFENYALAVDNG 120
DB 61 ELLYPTNASTLFLDIOEVGYVLIANQVRQVPLQRLIVRGTOLEFENYALAVDNG 120
QY 121 DPLNNTTPTVTSASPGRLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
DB 121 DPLNNTTPTVTSASPGRLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
QY 181 LTLIDPNRSRACHSPCKSGSRMGSESSDCCSLRTVAGGARGCKGLPTDCCHEOC 240
DB 181 LTLIDPNRSRACHSPCKSGSRMGSESSDCCSLRTVAGGARGCKGLPTDCCHEOC 240
QY 241 AAGCTGPRGSDCLALCFHNSGICELHCPALVTYNTDTPFSMENPEGRYTFGASCVTACP 300
DB 241 AAGCTGPRGSDCLALCFHNSGICELHCPALVTYNTDTPFSMENPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTIWCPLHNOEVTAEDETCQCEKSKCAVCYGLGMFNFTVSFWLR-V 359
DB 301 YNYLSTDVSGCTIWCPLHNOEVTAEDETCQCEKSKCAVCYGLGMFNFTVSFWLR-V 359
QY 353 301 YNYLSTDVSGCTIWCPLHNOEVTAEDETCQCEKSKCAVCYGLGMFNFTVSFWLR-V 359
DB 353 301 YNYLSTDVSGCTIWCPLHNOEVTAEDETCQCEKSKCAVCYGLGMFNFTVSFWLR-V 359
QY 360 PKVASHLSEFAGCKKIFGSLAFLESFSDGDPASNTAPLOPELOVYFETLEITGLYIS 419
DB 360 PKVASHLSEFAGCKKIFGSLAFLESFSDGDPASNTAPLOPELOVYFETLEITGLYIS 419
QY 419 360 PKVASHLSEFAGCKKIFGSLAFLESFSDGDPASNTAPLOPELOVYFETLEITGLYIS 419
DB 419 360 PKVASHLSEFAGCKKIFGSLAFLESFSDGDPASNTAPLOPELOVYFETLEITGLYIS 419
QY 420 AMPDGLPLSVFQWLQYIRGRILHNGAYSLTLOGLGISWGLSLRELGSGLALIHNNTH 479
DB 420 AMPDGLPLSVFQWLQYIRGRILHNGAYSLTLOGLGISWGLSLRELGSGLALIHNNTH 479
QY 479 420 AMPDGLPLSVFQWLQYIRGRILHNGAYSLTLOGLGISWGLSLRELGSGLALIHNNTH 479
DB 479 420 AMPDGLPLSVFQWLQYIRGRILHNGAYSLTLOGLGISWGLSLRELGSGLALIHNNTH 479
QY 480 LCFVHTVPMQDLFNNPQALLHTANREDECVGEGLAHQDLARGHGMGPGPOCNCSQ 539
DB 480 LCFVHTVPMQDLFNNPQALLHTANREDECVGEGLAHQDLARGHGMGPGPOCNCSQ 539
QY 539 480 LCFVHTVPMQDLFNNPQALLHTANREDECVGEGLAHQDLARGHGMGPGPOCNCSQ 539
DB 539 480 LCFVHTVPMQDLFNNPQALLHTANREDECVGEGLAHQDLARGHGMGPGPOCNCSQ 539
QY 540 FLRGQECVEGCVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKXP 599
DB 540 FLRGQECVEGCVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKXP 599
QY 599 540 FLRGQECVEGCVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKXP 599
DB 599 540 FLRGQECVEGCVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKXP 599
QY 600 PFCVARGSGVKKPLSLYMPIMKFPDEGACQPCPINCTHSCVULDXKGPABRASPILTS 659
DB 600 PFCVARGSGVKKPLSLYMPIMKFPDEGACQPCPINCTHSCVULDXKGPABRASPILTS 659
QY 659 600 PFCVARGSGVKKPLSLYMPIMKFPDEGACQPCPINCTHSCVULDXKGPABRASPILTS 659
DB 659 600 PFCVARGSGVKKPLSLYMPIMKFPDEGACQPCPINCTHSCVULDXKGPABRASPILTS 659

```

QY 660 IVSAVVGILLVVVLGVFGILLIKRROOKIRKXTYMRLLQETELVEPLTPSGAMPNOAOMR 719
 DB 654 IVSAVVGILLVVVLGVFGILLIKRROOKIRKXTYMRLLQETELVEPLTPSGAMPNOAOMR 713
 QY 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779
 DB 714 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 773
 QY 780 MAGVSPVYSRLIGLITSTVOLVTQMLPYGCLLHVRENRGLSGLIMCMQIAOMR 839
 DB 774 MAGVSPVYSRLIGLITSTVOLVTQMLPYGCLLHVRENRGLSGLIMCMQIAOMR 833
 QY 840 SYLEDVLYHRDLAARNVLYKSPNHVKITDFGLARLLIDETEVHADGKXPIKMMALES 899
 DB 834 SYLEDVLYHRDLAARNVLYKSPNHVKITDFGLARLLIDETEVHADGKXPIKMMALES 893
 QY 900 ILRRRFTHSDVMSXGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
 DB 894 ILRRRFTHSDVMSXGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
 QY 960 IMVKCMWIDSECRPRRELIVSEFSRWARDPQRFVYIIONEDLGASPLDSTFYRSLLEDD 1019
 DB 954 IMVKCMWIDSECRPRRELIVSEFSRWARDPQRFVYIIONEDLGASPLDSTFYRSLLEDD 1013
 QY 1020 MSDLVDAEEYLVVQOGFCPPDPAPGAGGMYHRHSSSTRSGGDLTIGLBPSEEARPS 1079
 DB 1014 MSDLVDAEEYLVVQOGFCPPDPAPGAGGMYHRHSSSTRSGGDLTIGLBPSEEARPS 1073
 QY 1080 PLAPSEGAGSDVDGDLGMAAKGLSLPTHPSPQRYSEDPVPLPSETDGYAPLTC 1139
 DB 1074 PLAPSEGAGSDVDGDLGMAAKGLSLPTHPSPQRYSEDPVPLPSETDGYAPLTC 1133
 QY 1140 SPQPEYVNOQDVPRPPSPREGPLPAAPAGATLEBAKTLSPGXKGVYKDYAFAGAVEN 1199
 DB 1134 SPQPEYVNOQDVPRPPSPREGPLPAAPAGATLEBAKTLSPGXKGVYKDYAFAGAVEN 1193
 QY 1200 PEYLTPOGGAAPRPPPAFSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYLGLDV 1259
 DB 1194 PEYLTPOGGAAPRPPPAFSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYLGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 3
 AAB60167
 ID AAB60167 standard; Protein; 1255 AA.
 XX AAB60167;
 AC AAB60167;
 DT 03-APR-2001 (first entry)
 XX
 DE HRR2 transgene plasmid construct encoded protein.
 XX
 KM Human; HRR2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 XX antibody.
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WC200100244-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000MO-US17229.
 XX
 PR 25-JUN-1999; 99US-0141316.
 XX 16-MAR-2000; 2000US-0189844.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Erickson, S, Schwall R;

XX WP1: 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX
 PT Treating tumors, particularly breast cancers, which overexpress an ErbB
 receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -
 XX
 PS Example 3; Fig 4; 92pp; English.
 XX
 CC The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.3%; Score 6736; DB 22; Length 1255;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 2;
 QY 1 MELAAICRWGILLALLPFGASTQVCTGDMKILRLPASPEFHLMRLHYOGCCOVQGNL 60
 DB 1 MELAAICRWGILLALLPFGASTQVCTGDMKILRLPASPEFHLMRLHYOGCCOVQGNL 60
 QY 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLBELQRLSTELIKGGLVLIQRPOLCYODTILMKDIFHKNOOLA 180
 DB 121 DPLNNTPTVTGASPGGLBELQRLSTELIKGGLVLIQRPOLCYODTILMKDIFHKNOOLA 180
 QY 181 LTLIDTNRBRACHPSGPKSGRSGESSEDCOSLTRVACGACRCKGRLPTCCHEQC 240
 DB 181 LTLIDTNRBRACHPSGPKSGRSGESSEDCOSLTRVACGACRCKGRLPTCCHEQC 240
 QY 241 AAGCTGPKASDCLACLFHNSGICELHCPALVYNTDTFESMPNBEGRYTGASCVTACP 300
 DB 241 AAGCTGPKASDCLACLFHNSGICELHCPALVYNTDTFESMPNBEGRYTGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRECKSKPCARVCYGLGMEH-----LREV 353
 DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRECKSKPCARVCYGLGMEH-----LREV 353
 QY 360 PKVASHLEEFAGCKKIFGSLAFLEPESFGDPASNTAPLQPEQLQVFETLEITGYLYIS 419
 DB 354 RAVTSANIOEFAGCKKIFGSLAFLEPESFGDPASNTAPLQPEQLQVFETLEITGYLYIS 413
 QY 420 AMPDSLPLDSVFQNLQVIRGRIUNGAVSLLTQGLGISWLGRLSRLGSLALIHNTH 479
 DB 414 AMPDSLPLDSVFQNLQVIRGRIUNGAVSLLTQGLGISWLGRLSRLGSLALIHNTH 473
 QY 480 LCFVHVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMPPTCCVNCSC 539
 DB 474 LCFVHVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMPPTCCVNCSC 533
 QY 540 FLRGQECVEECRVLQGLPREVYNABHCLPCHBECQOPNGSVTCFGEPEADQCVAACHYKDP 599
 DB 534 FLRGQECVEECRVLQGLPREVYNABHCLPCHBECQOPNGSVTCFGEPEADQCVAACHYKDP 593
 QY 600 PFCVAPCPGSGVPRDLSYPIWKFPPBEBACQPCPINCNSCVLDLDDKCPAEQASPLTS 659
 DB 594 PFCVAPCPGSGVPRDLSYPIWKFPPBEBACQPCPINCNSCVLDLDDKCPAEQASPLTS 653
 QY 660 IVSAVVGILLVVVLGVFGILLIKRROOKIRKXTYMRLLQETELVEPLTPSGAMPNOAOMR 719
 DB 654 IVSAVVGILLVVVLGVFGILLIKRROOKIRKXTYMRLLQETELVEPLTPSGAMPNOAOMR 713
 QY 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779

Db 714 ILKTELKRVKVLGSGAGFTYKGIWIDGENVKI PVAIKYLRNTSPKANKELIDEAYV 773
 Qy 780 MAGVSPYVSRLLGICLTSTVQLVTLQMLPYGCLDHEVNRGLRSGODLWMCQIAKGM 839
 Db 774 MAGVSPYVSRLLGICLTSTVQLVTLQMLPYGCLDHEVNRGLRSGODLWMCQIAKGM 833
 Qy 840 STLEDVRLVHRDLAARNVLVSPNHWKTTDGLARLLDIDETEHADGKVPIMMALES 899
 Db 834 STLEDVRLVHRDLAARNVLVSPNHWKTTDGLARLLDIDETEHADGKVPIMMALES 893
 Qy 900 ILRRFTHQSDVWSGVTWELMTGAKPYDGIAPAREIPDLLEXGERLPPOPICTIDVYM 959
 Db 894 ILRRFTHQSDVWSGVTWELMTGAKPYDGIAPAREIPDLLEXGERLPPOPICTIDVYM 953
 Qy 960 IWKCMWIDSECRPRPRELIVSEFSMARDPQRFVVIQNEGLGPAAPLDSITFRSLIEDDD 1019
 Db 954 IWKCMWIDSECRPRPRELIVSEFSMARDPQRFVVIQNEGLGPAAPLDSITFRSLIEDDD 1013
 Qy 1020 MGDLVDAEEYLVPOQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEARPS 1079
 Db 1014 MGDLVDAEEYLVPOQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEARPS 1073
 Qy 1080 PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTC 1139
 Db 1074 PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTC 1133
 Qy 1140 SPQPEYVNOPDVRPQPPSPRREGPLPAAPACATLBRATKLSPGKNGVYKDVFAFGAVEN 1199
 Db 1134 SPQPEYVNOPDVRPQPPSPRREGPLPAAPACATLBRATKLSPGKNGVYKDVFAFGAVEN 1193
 Qy 1200 PEYLTPOGGAAPQPPPPAFSPAFNLLYYMDODDPERGAPSTFKGPTAENPEYGLDY 1259
 Db 1194 PEYLTPOGGAAPQPPPPAFSPAFNLLYYMDODDPERGAPSTFKGPTAENPEYGLDY 1253
 Qy 1260 PY 1261
 Db 1254 PY 1255
 Db
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macropneagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US:002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOMSKI M.
 XX

PI Erickson S, Schwall R, Sliwowski M;
 XX MPI: 2002-163686/21.
 DR N-PSDB; ABR14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macropneagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 XX
 Query Match 98.3%; Score 6736; DB 23; Length 1255;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 1;
 Qy 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYOGCQVVOGNT 60
 Db 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYOGCQVVOGNT 60
 Qy 61 ELTYIPTNASISFLDIOEVGVYVLIHANOYRQVPLQRLAIYRGTOFLFEDNYALAVDNG 126
 Db 61 ELTYIPTNASISFLDIOEVGVYVLIHANOYRQVPLQRLAIYRGTOFLFEDNYALAVDNG 126
 Qy 121 DPLNNTTPTVGTASFGSLRELOLRSLTEILKSGVLIQRNPOLCYODTILMMDIFKKNQLA 186
 Db 121 DPLNNTTPTVGTASFGSLRELOLRSLTEILKSGVLIQRNPOLCYODTILMMDIFKKNQLA 186
 Qy 181 LTLIDTNSRACHPCSPCKGSRGWSSSEDCOSLRTVCAAGCARGKGPLPTDCHEQC 246
 Db 181 LTLIDTNSRACHPCSPCKGSRGWSSSEDCOSLRTVCAAGCARGKGPLPTDCHEQC 246
 Qy 241 AAGCTGPRGSDCLCLPHNNSGICELCPALVNTDTEFSMPBERYFFGASCVYACP 306
 Db 241 AAGCTGPRGSDCLCLPHNNSGICELCPALVNTDTEFSMPBERYFFGASCVYACP 306
 Qy 301 YNYLSTDVGSCTVLCPLHNEVTAEDGTORCEKSKPCARVCYGLGFMNFVFSFWLR-V 359
 Db 301 YNYLSTDVGSCTVLCPLHNEVTAEDGTORCEKSKPCARVCYGLGFMNFVFSFWLR-V 359
 Qy 360 PKVASHLIEFRAGCKKTFGSLAFLPSEFDGDPASNTAPLPDQQLVETLEETITGLYIS 419
 Db 360 PKVASHLIEFRAGCKKTFGSLAFLPSEFDGDPASNTAPLPDQQLVETLEETITGLYIS 419
 Qy 420 AMPDLPDLVSFYQMLQVIRGRIILHNGAYSLTLQGLISWLGRLSELGSLLIHHNTH 479
 Db 420 AMPDLPDLVSFYQMLQVIRGRIILHNGAYSLTLQGLISWLGRLSELGSLLIHHNTH 479
 Qy 480 LCFVHTVPMQDLFNNPQALLHTANREDECEVGEGLACHQLCARGCWGPGPTQVNCSCQ 539
 Db 480 LCFVHTVPMQDLFNNPQALLHTANREDECEVGEGLACHQLCARGCWGPGPTQVNCSCQ 539
 Qy 540 FLRGQCEVEECRYVQGLPREYVNAARHCLPCHPECOQNSVYTFGFEADQCAVACAYKXP 599
 Db 540 FLRGQCEVEECRYVQGLPREYVNAARHCLPCHPECOQNSVYTFGFEADQCAVACAYKXP 599
 Qy 600 PFVACRCPGKVPDLVSYPIMKPEDEGACQPCPINCTHSCVDLDDKCGPAEORASPLTS 659
 Db 600 PFVACRCPGKVPDLVSYPIMKPEDEGACQPCPINCTHSCVDLDDKCGPAEORASPLTS 659

```

Db      594 PFCVARSQGVKPDLSYBIMKFPDEEACQPCP.NCHTS.CVDLDDGCPCEQASLTS 653
Qy      660 IYSAVGILLVVLGVVGIILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAOMR 719
Db      654 IYSAVGILLVVLGVVGIILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAOMR 713
Qy      720 ILKETELRKVKVLSGAFGTYYKGIWIPDGENVKI.PVAIKVIRENTSPKANKEILDEAYV 779
Db      714 ILKETELRKVKVLSGAFGTYYKGIWIPDGENVKI.PVAIKVIRENTSPKANKEILDEAYV 773
Qy      780 MAGVSPVYSRLIGLITLTSTQVLTQVLMVPGGLDHNVENGRIGSDOLMWCQIAKGM 839
Db      774 MAGVSPVYSRLIGLITLTSTQVLTQVLMVPGGLDHNVENGRIGSDOLMWCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLKSPNHVKTIDFGLARLLIDETEVHADGKVP.IKMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLKSPNHVKTIDFGLARLLIDETEVHADGKVP.IKMMALES 893
Qy      900 ILRRRFTHSDVMSYGVYTWELMTFGAKYDGI.PAREIPDLLEKGERLPOPICITIDVYM 959
Db      894 ILRRRFTHSDVMSYGVYTWELMTFGAKYDGI.PAREIPDLLEKGERLPOPICITIDVYM 953
Qy      960 IMYKCMWIDSECPREELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDD 1019
Db      954 IMYKCMWIDSECPREELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEEYLVPOQGFPCDPAPAGAGVNHHRSSSTRSGGGLTLGLESESEARS 1079
Db      1014 MGDLYDAEEYLVPOQGFPCDPAPAGAGVNHHRSSSTRSGGGLTLGLESESEARS 1073
Qy      1080 PLASEGASDVDDGDLGMAAKGLQSLPTHPS.P.ORSSEDPVPLPSETDGYAAPLTC 1139
Db      1074 PLASEGASDVDDGDLGMAAKGLQSLPTHPS.P.ORSSEDPVPLPSETDGYAAPLTC 1133
Qy      1140 SPOPEYVNOCPVAPPPSPREGLPAPAPAGATLIERAKT.LSPQKGVVAVAFGAAVEN 1199
Db      1134 SPOPEYVNOCPVAPPPSPREGLPAPAPAGATLIERAKT.LSPQKGVVAVAFGAAVEN 1193
Qy      1200 PEYLTPGGGAAPQHPPAPSPAFDLVYMDOPPRERGA.PPSTFKGTPTAENPEYLGLDV 1259
Db      1194 PEYLTPGGGAAPQHPPAPSPAFDLVYMDOPPRERGA.PPSTFKGTPTAENPEYLGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 5
AA001111 standard; Protein; 1255 AA.
XX
AC AA001111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KM HER-2/neu: c-erbB1; p185; oncogene; tyrosine protein kinase;
KM breast cancer; ovary cancer; colon cancer; lung cancer;
KM prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
PH Key
FT Domain
FT Location/Qualifiers
FT /label= "intracellular domain"
FT /note= "claimed domain, useful for immunisation"
XX
XX WO9630514-A1.
XX
XX 03-OCT-1996.
XX

```

```

PF 28-MAR-1996; 96WO-US01689.
XX
XX 31-MAR-1995; 95US-0414417.
PR
XX (UNIM) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
PI
XX WPI, 1996-455361/45.
DR N-PSDB; AAT40739.
XX
XX DNA encoding HER-2-neu poly(peptide)s - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
XX Claim 2; Page 56-61; 71pp; English.
XX
XX Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
XX
Query Match 98.2%; Score 6730; DB 17; Length 1255;
Best local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2.
Qy 1 MELAAACRAGLLALLPAASTQVCTGDMKLR.PASPEHLDMLRLHLYGQCQVVGNTL 60
Db 1 MELAAACRAGLLALLPAASTQVCTGDMKLR.PASPEHLDMLRLHLYGQCQVVGNTL 60
Qy 61 ELTYLPNLSLFLQDIOGVGVILAHNQVQVPLORLIRYRGQLVEDNYALAVLDNG 120
Db 61 ELTYLPNLSLFLQDIOGVGVILAHNQVQVPLORLIRYRGQLVEDNYALAVLDNG 120
Qy 61 ELTYLPNLSLFLQDIOGVGVILAHNQVQVPLORLIRYRGQLVEDNYALAVLDNG 120
Db 61 ELTYLPNLSLFLQDIOGVGVILAHNQVQVPLORLIRYRGQLVEDNYALAVLDNG 120
Qy 121 DPLNNTPTVAGSPGGLRELQSLTEILKGVILQRPOLCYQDTILMKDIFHNQOLA 180
Db 121 DPLNNTPTVAGSPGGLRELQSLTEILKGVILQRPOLCYQDTILMKDIFHNQOLA 180
Qy 121 DPLNNTPTVAGSPGGLRELQSLTEILKGVILQRPOLCYQDTILMKDIFHNQOLA 180
Db 121 DPLNNTPTVAGSPGGLRELQSLTEILKGVILQRPOLCYQDTILMKDIFHNQOLA 180
Qy 181 LTLIDNRSRACHPCSPMCKGRS.CWSESSEDCQSLRTVACGACARCKGPLTDCCHQC 240
Db 181 LTLIDNRSRACHPCSPMCKGRS.CWSESSEDCQSLRTVACGACARCKGPLTDCCHQC 240
Qy 181 LTLIDNRSRACHPCSPMCKGRS.CWSESSEDCQSLRTVACGACARCKGPLTDCCHQC 240
Db 181 LTLIDNRSRACHPCSPMCKGRS.CWSESSEDCQSLRTVACGACARCKGPLTDCCHQC 240
Qy 241 AAGCTGKASDCLACLHPNHSICELHCAALTYNMDTFESMPNREGRTFGASCVTACP 300
Db 241 AAGCTGKASDCLACLHPNHSICELHCAALTYNMDTFESMPNREGRTFGASCVTACP 300
Qy 241 AAGCTGKASDCLACLHPNHSICELHCAALTYNMDTFESMPNREGRTFGASCVTACP 300
Db 241 AAGCTGKASDCLACLHPNHSICELHCAALTYNMDTFESMPNREGRTFGASCVTACP 300
Qy 301 YNYLSTVGSCTLVCP.LHNOEYVAEDGTORCEKSKPCARVCGYGLMFTVSFWLR-V 359
Db 301 YNYLSTVGSCTLVCP.LHNOEYVAEDGTORCEKSKPCARVCGYGLMFTVSFWLR-V 359
Qy 301 YNYLSTVGSCTLVCP.LHNOEYVAEDGTORCEKSKPCARVCGYGLMFTVSFWLR-V 359
Db 301 YNYLSTVGSCTLVCP.LHNOEYVAEDGTORCEKSKPCARVCGYGLMFTVSFWLR-V 359
Qy 360 PKVASHLPEFPAGCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 419
Db 360 PKVASHLPEFPAGCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 419
Qy 360 PKVASHLPEFPAGCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 419
Db 360 PKVASHLPEFPAGCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 419
Qy 354 RAYTSANIOEFAQCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 413
Db 354 RAYTSANIOEFAQCKIFGSLATLPSSFGDDPASNAPIQPELOVFEETLEITGLVYS 413
Qy 420 AMPSLDLSVFNQVLYRGRILHANGVSLTLOGLSIMLGLRSLRELQSGAL.IHHNTH 479
Db 420 AMPSLDLSVFNQVLYRGRILHANGVSLTLOGLSIMLGLRSLRELQSGAL.IHHNTH 479
Qy 420 AMPSLDLSVFNQVLYRGRILHANGVSLTLOGLSIMLGLRSLRELQSGAL.IHHNTH 479
Db 420 AMPSLDLSVFNQVLYRGRILHANGVSLTLOGLSIMLGLRSLRELQSGAL.IHHNTH 479
Qy 480 LCFVHTVPMQDLFRNPQALLHTANRDEDECEVGEGLAQQLCARHGCMWPGPTQVNCQ 539
Db 480 LCFVHTVPMQDLFRNPQALLHTANRDEDECEVGEGLAQQLCARHGCMWPGPTQVNCQ 539
Qy 474 LCFVHTVPMQDLFRNPQALLHTANRDEDECEVGEGLAQQLCARHGCMWPGPTQVNCQ 533
Db 474 LCFVHTVPMQDLFRNPQALLHTANRDEDECEVGEGLAQQLCARHGCMWPGPTQVNCQ 533
Qy 540 FLRGQCEVEGRVYVGLPREYVNAARHCLPCHPCOPONGSVTCFGEAOCVACAHYKXP 599
Db 540 FLRGQCEVEGRVYVGLPREYVNAARHCLPCHPCOPONGSVTCFGEAOCVACAHYKXP 599
Qy 534 FLRGQCEVEGRVYVGLPREYVNAARHCLPCHPCOPONGSVTCFGEAOCVACAHYKXP 593
Db 534 FLRGQCEVEGRVYVGLPREYVNAARHCLPCHPCOPONGSVTCFGEAOCVACAHYKXP 593

```

QY 600 PFCVACPSGVKPDLSYMPIMKFPDEBACQCPINCHSCVDLDDKCCPAEQASPLTS 659
 DB 594 PFCVACPSGVKPDLSYMPIMKFPDEBACQCPINCHSCVDLDDKCCPAEQASPLTS 653
 QY 660 ISAVVGIILVVVLGVFGILLKRSQOKIRKTYMRLLOETELVEPLTPSAGMNOQMR 719
 DB 654 ISAVVGIILVVVLGVFGILLKRSQOKIRKTYMRLLOETELVEPLTPSAGMNOQMR 713
 QY 720 ILKETELRKVKYLGSAGFGTGYKGIIMIDGENVKIPVAIKVLRNTSPKANKELIDEAYV 779
 DB 714 ILKETELRKVKYLGSAGFGTGYKGIIMIDGENVKIPVAIKVLRNTSPKANKELIDEAYV 773
 QY 780 MAGVSPVYSLILGICLTSTQVLTQVLMVYGLDHNENGRGLSQDLMLWCMQIAKGM 839
 DB 774 MAGVSPVYSLILGICLTSTQVLTQVLMVYGLDHNENGRGLSQDLMLWCMQIAKGM 833
 QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGKYPIMMALES 899
 DB 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGKYPIMMALES 893
 QY 900 ILRRRTHOSDVMSYGVYTWELMTGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
 DB 894 ILRRRTHOSDVMSYGVYTWELMTGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
 QY 960 IMVKCMIDSECRPRRELVSFESRMARDPQRFVYIOWEDIGPASPILDSTFYRSLLEDD 1019
 DB 954 IMVKCMIDSECRPRRELVSFESRMARDPQRFVYIOWEDIGPASPILDSTFYRSLLEDD 1013
 QY 1020 MGDLYDAEEYLVPOQGFPCPDPAFAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRS 1079
 DB 1014 MGDLYDAEEYLVPOQGFPCPDPAFAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRS 1073
 QY 1080 PLAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSSDPVPLPSEITDGYVAELTC 1139
 DB 1074 PLAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSSDPVPLPSEITDGYVAELTC 1133
 QY 1140 SPQPEYVNPQDVROPSPREBGLPAARPAATLERAKTSLPGKGVYKDVFAFGAVEN 1199
 DB 1134 SPQPEYVNPQDVROPSPREBGLPAARPAATLERAKTSLPGKGVYKDVFAFGAVEN 1193
 QY 1200 PEYLPQGGAAPOHPHPPAFSAFNNLYYMODPPERGAPOSTKGTPTANPEYLGLDV 1259
 DB 1194 PEYLPQGGAAPOHPHPPAFSAFNNLYYMODPPERGAPOSTKGTPTANPEYLGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 6
 AAM92406
 ID AAM92406 standard; Protein; 1255 AA.
 AC
 XX AAM92406;
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 DE
 KM HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KM malignancy; treatment; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 676..1255
 FT /note="region which elicits immune response"
 XX
 PN US5869445-A.
 XX
 XX 09-FEB-1999.
 PD
 XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 PA (UNIV) UNIV WASHINGTON.
 XX
 PI Cheever MA, Disle ML;
 XX
 DR WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 PS Claim 3; Column 31-38; 26pp; English.
 XX
 CC This sequence represents the human HER-2/neu oncogene protein. A fragmen
 CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 CC
 XX
 SQ Sequence 1255 AA;
 Query Match 98.2%; Score 6730; DB 20; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0; Mismatches 4; Indels 8; Gaps
 Matches 1242; Conservative 8; N-matches 4; Indels 8; Gaps 8;
 QY 1 METALCRWGLLALLPFGAASVQCTGTDKRLPASPETHLMDRLHYGCGVQGNL 60
 DB 1 METALCRWGLLALLPFGAASVQCTGTDKRLPASPETHLMDRLHYGCGVQGNL 60
 QY 61 ELYLPPTASLSFLDIDQEVGVYLIANNQVQVPLQRLVRSQTQLFEDNYALAVDNG 121
 DB 61 ELYLPPTASLSFLDIDQEVGVYLIANNQVQVPLQRLVRSQTQLFEDNYALAVDNG 121
 QY 121 DELNNTTPTVGTASPGGLRELOLRSLTEILKGVYIQRNPOLCYQDTIMKDIPIKNNOLA 181
 DB 121 DELNNTTPTVGTASPGGLRELOLRSLTEILKGVYIQRNPOLCYQDTIMKDIPIKNNOLA 181
 QY 181 LTLIDTNSRACHPSCPMCKSGRCKGSESSDQSLTRTVACAGCARCKGRLPTDCHEQC 241
 DB 181 LTLIDTNSRACHPSCPMCKSGRCKGSESSDQSLTRTVACAGCARCKGRLPTDCHEQC 241
 QY 241 AAGCTGPKHSDCLALHFNHSGICEIHPALVYNTPTFESMPNPEGRYTFGASCCTACP 301
 DB 241 AAGCTGPKHSDCLALHFNHSGICEIHPALVYNTPTFESMPNPEGRYTFGASCCTACP 301
 QY 301 YNYLSTDVSGCTLVCPHNGEVTAEEDGTQRCCKSKPCARVCYGLGMEH-----LREV 351
 DB 301 YNYLSTDVSGCTLVCPHNGEVTAEEDGTQRCCKSKPCARVCYGLGMEH-----LREV 351
 QY 351 YNYLSTDVSGCTLVCPHNGEVTAEEDGTQRCCKSKPCARVCYGLGMEH-----LREV 351
 DB 351 YNYLSTDVSGCTLVCPHNGEVTAEEDGTQRCCKSKPCARVCYGLGMEH-----LREV 351
 QY 360 PVSASHLEEFAGCKTIGSLAPFPESDGPASNTAPLOEQVQVETLEITGYLYIS 415
 DB 360 PVSASHLEEFAGCKTIGSLAPFPESDGPASNTAPLOEQVQVETLEITGYLYIS 415
 QY 415 PVSASHLEEFAGCKTIGSLAPFPESDGPASNTAPLOEQVQVETLEITGYLYIS 415
 DB 415 PVSASHLEEFAGCKTIGSLAPFPESDGPASNTAPLOEQVQVETLEITGYLYIS 415
 QY 420 AMPDSLPLSVFONLOVIRGRILHNGAVSLTLQGLISMLGLSLRLSGSLALIHNTH 475
 DB 420 AMPDSLPLSVFONLOVIRGRILHNGAVSLTLQGLISMLGLSLRLSGSLALIHNTH 475
 QY 475 LCFYHTVWMDQLFRNPQALHTANREDEDEVGGLACHOLCARGHCMGPGPTQCVNCSCQ 535
 DB 475 LCFYHTVWMDQLFRNPQALHTANREDEDEVGGLACHOLCARGHCMGPGPTQCVNCSCQ 535
 QY 540 FLRGQCEVCECRVQLGPREVYNARHCLPCHPEQOPNGSVTGGPZADQVCAHAKDP 599
 DB 540 FLRGQCEVCECRVQLGPREVYNARHCLPCHPEQOPNGSVTGGPZADQVCAHAKDP 599

```

QY 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQCPICNTHSCVDLDKCGAPAEORASPLTS 659
DB 594 PFCVACPSGVKPDLSYMPIMKFPDEGACQCPICNTHSCVDLDKCGAPAEORASPLTS 653
QY 660 TUSAIVGILLVTVLVGVFGLIKRQOKRKTYMRLLQETELVEPLTPSGAMPNOQMR 719
DB 654 IISAIVGILLVTVLVGVFGLIKRQOKRKTYMRLLQETELVEPLTPSGAMPNOQMR 713
QY 720 ILKETELRKVKVLGSGAFGVYKGIIPGENVKIPVAIKVLRNTSPKANKELIDEYV 779
DB 714 ILKETELRKVKVLGSGAFGVYKGIIPGENVKIPVAIKVLRNTSPKANKELIDEYV 773
QY 780 MAGVSPYVSRLLIGILCTSTVQLVTLQMPYGLLDHVRENRGLSGDILMCMQIAKM 839
DB 774 MAGVSPYVSRLLIGILCTSTVQLVTLQMPYGLLDHVRENRGLSGDILMCMQIAKM 833
QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKINDPGLALDIDETEHADGKXPIKMALES 899
DB 834 SYLEDVRLVHRDLAARNVLKSPNHVKINDPGLALDIDETEHADGKXPIKMALES 893
QY 900 ILRRFTHQS DVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRFTHQS DVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVKCWMIDSECHPRRELVSFSRMAARDPQRFVVIQNEIDGFPASPLDSTFYRSLLEDD 1019
DB 954 IMVKCWMIDSECHPRRELVSFSRMAARDPQRFVVIQNEIDGFPASPLDSTFYRSLLEDD 1013
QY 1020 MGDVDAEEYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLPSEEAAPS 1079
DB 1014 MGDVDAEEYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLPSEEAAPS 1073
QY 1080 PLAPSEGASDVFDGDLGMAKGLQSLPTHDPSELQRYSEDPVPLPSETDGYVAPLTC 1139
DB 1074 PLAPSEGASDVFDGDLGMAKGLQSLPTHDPSELQRYSEDPVPLPSETDGYVAPLTC 1133
QY 1140 SPOPEYVNOPDVRPOPSPREGPLPAAPAGATLERAKTLSGKXGVKDVFAFGAVEN 1199
DB 1134 SPOPEYVNOPDVRPOPSPREGPLPAAPAGATLERAKTLSGKXGVKDVFAFGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPAPFSPAFNLVYWDQDPBERGAPSTFGKPTAENPEYLGIDV 1259
DB 1194 PEYLTPOGGAAPQHPAPFSPAFNLVYWDQDPBERGAPSTFGKPTAENPEYLGIDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

```

RESULT 7
AAB21198
ID AAB21198 standard: protein; 1255 AA.

XX AAB21198;
XX 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
XX Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer.
XX Homo sapiens.
XX OS
XX PN MO2000044899-A1.
XX PD 03-AUG-2000.
XX 28-JAN-2000; 2000MO-US02164.
XX PF 29-JAN-1999; 99US-0117976.
XX PR

```

PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
PI Cheever MA, Gheylen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers.
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;
QY 1 MELAALCRKGLLALPPGAASVQVCTGDMKRLPASPTHLMRLHYOGCOWQGNL 60
DB 1 MELAALCRKGLLALPPGAASVQVCTGDMKRLPASPTHLMRLHYOGCOWQGNL 60
QY 61 ELITYPTNAGSLFLODIEVOGVYIIAHNVQVPLQRLIVRGTOLEFEDNYALAVLDNG 120
DB 61 ELITYPTNAGSLFLODIEVOGVYIIAHNVQVPLQRLIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVAGSPGGLRELOKRSITELIKGGVLIQRNPOLCYODTILMKOIFKXNOLA 180
DB 121 DPLNNTPTVAGSPGGLRELOKRSITELIKGGVLIQRNPOLCYODTILMKOIFKXNOLA 180
QY 181 LTLIDNRSRACHPCSPKCKSGRSGESSDQSLTRVACAGCARCKGSLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPKCKSGRSGESSDQSLTRVACAGCARCKGSLPTDCHEQC 240
QY 241 AAGCTGPRHSDCLALHNHSGICELHPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AAGCTGPRHSDCLALHNHSGICELHPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVSGCTVLCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGFMNNFTVSFMLR-V 354
DB 301 YNYLSTDVSGCTVLCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGFMNNFTVSFMLR-V 354
QY 354 RAVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 413
DB 354 RAVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 413
QY 420 AWPDSLPLSVFQNLQVRRGRIILHNGAVSLTLQGLGSMIGLSRLRELSGLALIHNNTH 475
DB 420 AWPDSLPLSVFQNLQVRRGRIILHNGAVSLTLQGLGSMIGLSRLRELSGLALIHNNTH 475
QY 475 AWPDSLPLSVFQNLQVRRGRIILHNGAVSLTLQGLGSMIGLSRLRELSGLALIHNNTH 475
DB 475 AWPDSLPLSVFQNLQVRRGRIILHNGAVSLTLQGLGSMIGLSRLRELSGLALIHNNTH 475
QY 480 LCFVHTVPMQDLFRNPQALHTANREDECVGEGALCHQLCARGHCGMGPTQCVNCSQ 539
DB 480 LCFVHTVPMQDLFRNPQALHTANREDECVGEGALCHQLCARGHCGMGPTQCVNCSQ 539
QY 539 LCFVHTVPMQDLFRNPQALHTANREDECVGEGALCHQLCARGHCGMGPTQCVNCSQ 539
DB 539 LCFVHTVPMQDLFRNPQALHTANREDECVGEGALCHQLCARGHCGMGPTQCVNCSQ 539
QY 540 FLRGOECVBEGRVIOGUPREYVNAHCLPCHPBCQPONGSVTTCGPPADQCVACAHYKDP 599
DB 540 FLRGOECVBEGRVIOGUPREYVNAHCLPCHPBCQPONGSVTTCGPPADQCVACAHYKDP 599
QY 599 FLRGOECVBEGRVIOGUPREYVNAHCLPCHPBCQPONGSVTTCGPPADQCVACAHYKDP 599
DB 599 FLRGOECVBEGRVIOGUPREYVNAHCLPCHPBCQPONGSVTTCGPPADQCVACAHYKDP 599
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQCPICNTHSCVDLDKCGAPAEORASPLTS 659
DB 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQCPICNTHSCVDLDKCGAPAEORASPLTS 659

```

```

Db      554 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVLDLDDGCCAEPORASPLTS 653
Qy      660 IVSAVVGILLVVLGVVFGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAKR 719
Db      654 IISAVVGILLVVLGVVFGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAKR 713
Qy      720 ILKETELRKVKVLGSGAGFGTYKGIWTDGNVNI PVAIKYLRNTSPKANKETLDEAYV 779
Db      714 ILKETELRKVKVLGSGAGFGTYKGIWTDGNVNI PVAIKYLRNTSPKANKETLDEAYV 773
Qy      780 MAGVSPYVSRLLGICTSTVOLVTQMPYCCLDHYRENGRGRSGODLNMCKQIAKGM 839
Db      774 MAGVSPYVSRLLGICTSTVOLVTQMPYCCLDHYRENGRGRSGODLNMCKQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKMMALES 893
Qy      900 ILRRFTHQSDVWSYGVTVWELMTFGAKPYGIPAREIPDLLEKGEPLPQPICTIDVTM 959
Db      894 ILRRFTHQSDVWSYGVTVWELMTFGAKPYGIPAREIPDLLEKGEPLPQPICTIDVTM 953
Qy      960 IMVKCMIDSECRPFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMIDSECRPFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEEYLVPOQGFPCPPAGAGAGMHHRRSSSTSGGDLTLGLEPSEEBAPRS 1079
Db      1014 MGDLYDAEEYLVPOQGFPCPPAGAGAGMHHRRSSSTSGGDLTLGLEPSEEBAPRS 1073
Qy      1080 PLAPSEGASDVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDYVAPLTC 1139
Db      1074 PLAPSEGASDVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDYVAPLTC 1133
Qy      1140 SPOPEYVNPQPVVRQPPSPREGRPLAARPAAGATLERKTLSPKNGVYKDVAFGAVEN 1199
Db      1134 SPOPEYVNPQPVVRQPPSPREGRPLAARPAAGATLERKTLSPKNGVYKDVAFGAVEN 1193
Qy      1200 PEYLTPQGAAPQHPPPAPSPADNLYYWDOPPEPAGPSTFKGPTAENPEYLGIDV 1259
Db      1194 PEYLTPQGAAPQHPPPAPSPADNLYYWDOPPEPAGPSTFKGPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AA#84780
ID      AA#84780 standard; Protein; 1255 AA.
AC      AA#84780;
DT      08-AUG-2000 (first entry)
XX      Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX      SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
XX      tumor cell proliferation; tissue degeneration; arthropathy;
XX      bone resorption; inflammatory disease; degenerative disorder;
XX      wound healing.
XX      Homo sapiens.
XX      OS
XX      PN      W0200020579-A1.
XX      PD      13-APR-2000.
XX      PE      01-OCT-1999; 99MO-CA00912.
XX      PR      02-OCT-1998; 98US-0165192.
XX

```

```

PA      (UYMC-) UNIV MCMMASTER.
XX      Muller WJ, Siegel PM;
XX      WPI. 2000-303768/26.
DR      N-PSDB; AAA14612.
XX
XX      Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX      erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erbB-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbB-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA:
Query Match      98.2%; Score 6730; DB 21; Length 1255;
Best Local Similarity 98.4%; Pred. No 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy      1 MELALCRWGLLLALPPGAASIQVCTGDMKRLPASPETHLDMRLHYQGQVQVGNL 60
Db      1 MELALCRWGLLLALPPGAASIQVCTGDMKRLPASPETHLDMRLHYQGQVQVGNL 60
Qy      61 ELTYLPTNASISFLODIOEVQGVVLAHNOVROVQRLRIYRGTLQFEDNYALAVLDNG 120
Db      61 ELTYLPTNASISFLODIOEVQGVVLAHNOVROVQRLRIYRGTLQFEDNYALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGRLRELQLRLSTLEILKGVLLIQRNPOLCYQDTILMKDIFHKANQLA 180
Db      121 DPLNNTTPVTGASPGRLRELQLRLSTLEILKGVLLIQRNPOLCYQDTILMKDIFHKANQLA 180
Qy      121 DPLNNTTPVTGASPGRLRELQLRLSTLEILKGVLLIQRNPOLCYQDTILMKDIFHKANQLA 180
Db      121 DPLNNTTPVTGASPGRLRELQLRLSTLEILKGVLLIQRNPOLCYQDTILMKDIFHKANQLA 180
Qy      181 LTLIDNTRSAACHPCSPMCKSGSKWGESSEDCSLRTYVQAGGACRCKPLPTDCHEQC 240
Db      181 LTLIDNTRSAACHPCSPMCKSGSKWGESSEDCSLRTYVQAGGACRCKPLPTDCHEQC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICEJHCPALVYNTDFESMNPREGRTFPASCVTACP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICEJHCPALVYNTDFESMNPREGRTFPASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMFNNFTVSEMLR-V 359
Db      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMFNNFTVSEMLR-V 353
Qy      360 PKYSASHLEEFACCKXIFPSLAFPESPGDPASNAPIQFQLOVFFETLEITGYLYIS 419
Db      354 RAYTSANIQFACCKXIFPSLAFPESPGDPASNAPIQFQLOVFFETLEITGYLYIS 413
Qy      420 AMPDSLJDSLVFQNLQVIRGRLIHNGAVSLTQGLGISWLGRLSRLQSGALATHNHTH 479
Db      414 AMPDSLJDSLVFQNLQVIRGRLIHNGAVSLTQGLGISWLGRLSRLQSGALATHNHTH 473
Qy      480 LCFVHTVPMQDLFRNHQALHTARPEDECVCEGLAQCLCARGHCMGPGFTQCVCNSQ 539
Db      474 LCFVHTVPMQDLFRNHQALHTARPEDECVCEGLAQCLCARGHCMGPGFTQCVCNSQ 533
Qy      540 FLRGQCEVECEKRYLQGLPREYVNAHRCJLCHPEQCPQNGSVTCFPEADQVCACHYKDP 599
Db      534 FLRGQCEVECEKRYLQGLPREYVNAHRCJLCHPEQCPQNGSVTCFPEADQVCACHYKDP 593
Qy      600 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVLDLDDGCCAEPORASPLTS 659

```

```

Db      594 PFCVARCPSGVKEDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQASPLTS 653
Qy      660 IYSAVVGILLVWVGVGVFGILIRKROKIRKYMRLLOETELVEBLTPSGAMPQAOQR 719
Db      654 IISAVVGILLVWVGVFGILIRKROKIRKYMRLLOETELVEBLTPSGAMPQAOQR 713
Qy      720 ILKETELRKVKVIGSGAFGTIVYKGIWIPDGENYKIPVAKIVLENTSPANKELIDEAVY 779
Db      714 ILKETELRKVKVIGSGAFGTIVYKGIWIPDGENYKIPVAKIVLENTSPANKELIDEAVY 773
Qy      760 MAGVSPVYSRLIGICLTSTVOLTOLMPYGCCLDHYRENRKGLSGODLLNMCQIAKQM 839
Db      774 MAGVSPVYSRLIGICLTSTVOLTOLMPYGCCLDHYRENRKGLSGODLLNMCQIAKQM 833
Qy      840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMALES 899
Db      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMALES 893
Qy      900 ILRRFTHQSDVMSYGVTVWELMTFEGAKPYDGIIPAREIPDLLEKGRLPQPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFEGAKPYDGIIPAREIPDLLEKGRLPQPICTIDVYM 953
Qy      960 IMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEYLVPOQGFCDPPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEPAPS 1079
Db      1014 MGDLYDAEYLVPOQGFCDPPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEPAPS 1073
Qy      1080 PLAPSEGAGSDVFDGDLGMAAGKLGSLPTHPDPSPLQRYSEDPVPLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMAAGKLGSLPTHPDPSPLQRYSEDPVPLPSETDGYVAPLTC 1133
Qy      1140 SPQPEYVNPQDVPRPQPSPREGPIPAARPAATLERKATISPGKNGVYKVFAGAVEN 1199
Db      1134 SPQPEYVNPQDVPRPQPSPREGPIPAARPAATLERKATISPGKNGVYKVFAGAVEN 1193
Qy      1200 PEYLPQGAAPQPPPAFSPAFENDLYWQDPPERGAAPPSTFKGTPTAENPEYLGLDV 1259
Db      1194 PEYLPQGAAPQPPPAFSPAFENDLYWQDPPERGAAPPSTFKGTPTAENPEYLGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytoskeletal; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN MO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (COR1-) COR1XA CORP.
XX

```

```

PI Cheever MA, Hand-Zimmermann S;
XX WPI, 2001-476112/51.
DR N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2, Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
XX
Query Match 98.2%; Score 6730; DB 22; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy 1 MELALCRWGLLALPPGAASQVCTGDMKRLPASPETHLMDRLHYOGCQVVOGNTL 60
Db 1 MELALCRWGLLALPPGAASQVCTGDMKRLPASPETHLMDRLHYOGCQVVOGNTL 60
Qy 61 ELTYLPTNASISFLQDIQEVQGVYVLIHANOQVQVPIQRLRIYVGTOLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASISFLQDIQEVQGVYVLIHANOQVQVPIQRLRIYVGTOLFEDNVALAVLDNG 120
Qy 121 DPLNNTPTVYGASGGVRELQRLSLTEILKGVLIORNPOLCTQDTITLWMDIFKKNQOLA 180
Db 121 DPLNNTPTVYGASGGVRELQRLSLTEILKGVLIORNPOLCTQDTITLWMDIFKKNQOLA 180
Qy 121 DPLNNTPTVYGASGGVRELQRLSLTEILKGVLIORNPOLCTQDTITLWMDIFKKNQOLA 180
Db 121 DPLNNTPTVYGASGGVRELQRLSLTEILKGVLIORNPOLCTQDTITLWMDIFKKNQOLA 180
Qy 161 LTLIDYNSRACHPCSPWCKSGRSCSSSDQSLRTVQAGGACARCKGPLPDCHEQC 240
Db 161 LTLIDYNSRACHPCSPWCKSGRSCSSSDQSLRTVQAGGACARCKGPLPDCHEQC 240
Qy 181 LTLIDYNSRACHPCSPWCKSGRSCSSSDQSLRTVQAGGACARCKGPLPDCHEQC 240
Db 181 LTLIDYNSRACHPCSPWCKSGRSCSSSDQSLRTVQAGGACARCKGPLPDCHEQC 240
Qy 241 AAGCTGPRHSQCLACIPHNSTGICELHCPALMYNTDFFESMNPBGRYFSGACTAC 300
Db 241 AAGCTGPRHSQCLACIPHNSTGICELHCPALMYNTDFFESMNPBGRYFSGACTAC 300
Qy 241 AAGCTGPRHSQCLACIPHNSTGICELHCPALMYNTDFFESMNPBGRYFSGACTAC 300
Db 241 AAGCTGPRHSQCLACIPHNSTGICELHCPALMYNTDFFESMNPBGRYFSGACTAC 300
Qy 301 YNYLSTDVGSCTIVCPHNOEVTAEQTCRCEKSKPCARVYCYGLGMFNFTVSFWLR-V 359
Db 301 YNYLSTDVGSCTIVCPHNOEVTAEQTCRCEKSKPCARVYCYGLGMFNFTVSFWLR-V 359
Qy 360 PKVSAHLSEFPAGCKKIFGSLAF.PESFDQDPASNTAPLOPELOQVETLEETIGVLYIS 419
Db 360 PKVSAHLSEFPAGCKKIFGSLAF.PESFDQDPASNTAPLOPELOQVETLEETIGVLYIS 419
Qy 354 RAVTSANIQEFPAGCKKIFGSLAF.PESFDQDPASNTAPLOPELOQVETLEETIGVLYIS 413
Db 354 RAVTSANIQEFPAGCKKIFGSLAF.PESFDQDPASNTAPLOPELOQVETLEETIGVLYIS 413
Qy 420 AMPDLPDLVSFQNLQVIRGRILANGAYSLTLQGLISWGLSLRELISGLALIHNTHT 479
Db 420 AMPDLPDLVSFQNLQVIRGRILANGAYSLTLQGLISWGLSLRELISGLALIHNTHT 479
Qy 414 AMPDLPDLVSFQNLQVIRGRILANGAYSLTLQGLISWGLSLRELISGLALIHNTHT 473
Db 414 AMPDLPDLVSFQNLQVIRGRILANGAYSLTLQGLISWGLSLRELISGLALIHNTHT 473
Qy 480 LCFVHTVPMQDLFNNPHQALHTANREDECVSGGLACHQLCARGHWGSGFPQCVNCSQ 539
Db 480 LCFVHTVPMQDLFNNPHQALHTANREDECVSGGLACHQLCARGHWGSGFPQCVNCSQ 539
Qy 474 LCFVHTVPMQDLFNNPHQALHTANREDECVSGGLACHQLCARGHWGSGFPQCVNCSQ 533
Db 474 LCFVHTVPMQDLFNNPHQALHTANREDECVSGGLACHQLCARGHWGSGFPQCVNCSQ 533
Qy 540 FLRQGEVCEECRVLOGLPREYVYARHCIPCHPEQOPNGSVTCGFPADQVCAHYKDP 599
Db 540 FLRQGEVCEECRVLOGLPREYVYARHCIPCHPEQOPNGSVTCGFPADQVCAHYKDP 599
Qy 534 FLRQGEVCEECRVLOGLPREYVYARHCIPCHPEQOPNGSVTCGFPADQVCAHYKDP 593
Db 534 FLRQGEVCEECRVLOGLPREYVYARHCIPCHPEQOPNGSVTCGFPADQVCAHYKDP 593
Qy 600 PFCVARCPSGVKPDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQASPLTS 655
Db 594 PFCVARCPSGVKPDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQASPLTS 653
Qy 660 IYSAVVGILLVWVGVFGILIRKROKIRKYMRLLOETELVEBLTPSGAMPQAOQR 715
Db 654 IYSAVVGILLVWVGVFGILIRKROKIRKYMRLLOETELVEBLTPSGAMPQAOQR 713

```



```

Db      654 IISAVGILLVVLGVFGILLIKRROQIKRKYTMRLLOETELVEPLTPSGAMENQOMR 713
QY      720 IKETELRKVKVLGSGAGTGVYKGIIMPDGENVKIPVAIKYLRENTSPKANEILDAVY 779
Db      714 IKETELRKVKVLGSGAGTGVYKGIIMPDGENVKIPVAIKYLRENTSPKANEILDAVY 773
QY      780 MAGVSPVYSRLGLGICLTSTVOLTQMLPYGCLLDHVENRGRLGSDLLMCMQIAKM 839
Db      774 MAGVSPVYSRLGLGICLTSTVOLTQMLPYGCLLDHVENRGRLGSDLLMCMQIAKM 833
QY      840 SYLEDVRLVHRDLAARNLVKSPNHVKITDPGLARLIDIDETEVHADGGKPIKIMWALLES 899
Db      834 SYLEDVRLVHRDLAARNLVKSPNHVKITDPGLARLIDIDETEVHADGGKPIKIMWALLES 893
QY      900 IIRRFTHQSDVWSYGVTVWELMTFGAARPDYDIPAREIPDLLEKGERLPOPPICITIDVM 959
Db      894 IIRRFTHQSDVWSYGVTVWELMTFGAARPDYDIPAREIPDLLEKGERLPOPPICITIDVM 953
QY      960 IMVKCMIDSECRPRPRELVSFSPMARDPQRFVIONEDIGPASPLDSTTYRSLLEDD 1019
Db      954 IMVKCMIDSECRPRPRELVSFSPMARDPQRFVIONEDIGPASPLDSTTYRSLLEDD 1013
QY      1020 MGDLYDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRS 1079
Db      1014 MGDLYDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRS 1073
QY      1080 PLAPSEAGASDFDGDLMGAAGKGLQSLPTDPSFLQYSEDPTVPLPSETDGVAPLTC 1139
Db      1074 PLAPSEAGASDFDGDLMGAAGKGLQSLPTDPSFLQYSEDPTVPLPSETDGVAPLTC 1133
QY      1140 SPQPEYVNPQDVRPOPSPREGLPLAASPAATLERAKTILSPKXGVVKKVFAFGAVEN 1199
Db      1134 SPQPEYVNPQDVRPOPSPREGLPLAASPAATLERAKTILSPKXGVVKKVFAFGAVEN 1193
QY      1200 PEYLTPOGGAAPQHPPPAFSPAFNLYYWDQDPERGAPSTKGTPTAENPEYGLDV 1259
Db      1194 PEYLTPOGGAAPQHPPPAFSPAFNLYYWDQDPERGAPSTKGTPTAENPEYGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

```

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.

```

AC      AAG88267;
DT      11-SEP-2001 (first entry)
XX      HER2/neu amino acid sequence.
DE      HER2/neu amino acid sequence.
XX      Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX      Immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
XX      tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX      Homo sapiens.
XX      OS
XX      WO200141787-A1.
XX      PD
XX      14-JUN-2001.
XX      PF
XX      11-DEC-2000; 2000WO-US31591.
XX      PR
XX      10-DEC-1999; 99US-0458299.
XX      PA
XX      (EPI-M-) EPIMUNE INC.
XX      PI
XX      Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX      Keogh E;
XX      WPI, 2001-374995/39.

```

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer
XX
XX
PS Disclosure, Page 15, 199pp, English.

XX The present invention describes isolated prepared HER2/neu epitopes (1).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (i), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (ii) comprising (i
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu
CC (3) a vaccine composition (iii) comprising (ii) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (i); and (5) an isolated nucleic acid encoding (ii). (i) has cytotoxic
CC and immunostimulant activities, and can be used in vaccines. (i), (ii)
CC and (iii) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (i) and (ii) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (i) or (ii). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX

Sequence 1255 AA;

Query Match 98.2%; Score 6730; DB 22; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps :

```

QY      1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPEHLDMLNLYOGCGVOGNTL 60
T      1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPEHLDMLNLYOGCGVOGNTL 60
QY      61 ETTYPTAASLSFLDIOEVGVYLIANQVQVPLQRLIRVGTQVLFEDNYALAVLDNG 120
Db      61 ETTYPTAASLSFLDIOEVGVYLIANQVQVPLQRLIRVGTQVLFEDNYALAVLDNG 120
QY      121 DPLNNTTAVTGASPGGLREIQLRSILTEILKGVLIQRNPOLCYODTILMKDIFKXNOLA 180
Db      121 DPLNNTTAVTGASPGGLREIQLRSILTEILKGVLIQRNPOLCYODTILMKDIFKXNOLA 180
QY      181 LTLIDTNRSPACHPCSPKCKSRGSESSDQSLRTVCAGGACGKGLPTDCHEQC 240
Db      181 LTLIDTNRSPACHPCSPKCKSRGSESSDQSLRTVCAGGACGKGLPTDCHEQC 240
QY      241 AAGCGPAGSDCLALPHNHSIGELCPALVYNTDTEFSMPBERYFGASCVYACP 300
Db      241 AAGCGPAGSDCLALPHNHSIGELCPALVYNTDTEFSMPBERYFGASCVYACP 300
QY      301 YNYLSTDVGSCTLVCPJHNCDEVTAEQDTCRCKSPCARVCYGLGFMNFVFSFWLR-V 359
Db      301 YNYLSTDVGSCTLVCPJHNCDEVTAEQDTCRCKSPCARVCYGLGFMNFVFSFWLR-V 359
QY      360 PKVASHLFEFAGCKITFGSLAFIPSEFDGDPASNTAPLPBOLQVFTLEETITGLYIS 419
Db      360 PKVASHLFEFAGCKITFGSLAFIPSEFDGDPASNTAPLPBOLQVFTLEETITGLYIS 419
QY      420 AMPDLPDLISVFNQLQVIRGRILNNGAYSLTLOGLGSMGLRSLRELGGALLIHNTH 479
Db      420 AMPDLPDLISVFNQLQVIRGRILNNGAYSLTLOGLGSMGLRSLRELGGALLIHNTH 479
QY      480 LCFVHTVPMQDLFFNPQALLHTANREDECVBGLACQGLCARGHCWGPPTQCVNCSQ 539

```

```

Db      474 LCFVHTVPMDOQLFRNPHQALLHTANRPEDECVGEGLAHQCLCARGHGMGSPPTQCVNCSQ 533
Qy      540 FLRGOCEVEBCRVLQGLPREYVNAHCLPCHEPCOPONGSVTCFGEPAQCVAQAHYKDP 599
Db      534 FLRGOCEVEBCRVLQGLPREYVNAHCLPCHEPCOPONGSVTCFGEPAQCVAQAHYKDP 593
Qy      600 PCVAPRCSPSVKPLDLSYMPIKPFDEGACOPCINCTHSCVLDLDDGCAEGRASPUS 659
Db      594 PCVAPRCSPSVKPLDLSYMPIKPFDEGACOPCINCTHSCVLDLDDGCAEGRASPUS 653
Qy      660 IVSAVAVGILLVVLGVVFGLIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNOAKR 719
Db      654 IISAVVGLLVVLGVVFGLIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNOAKR 713
Qy      720 IUKETELKRVKVLGSGAGTGYKGIWIPDGNVYKIPAIKYLRENTSPKANKILLDEAY 779
Db      714 IUKETELKRVKVLGSGAGTGYKGIWIPDGNVYKIPAIKYLRENTSPKANKILLDEAY 773
Qy      780 MAGVSPYVSRLLGICLTSTVQLVTQLMPIYGLLDHVENRGRLGSODLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLLGICLTSTVQLVTQLMPIYGLLDHVENRGRLGSODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLYSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLYSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALES 893
Qy      900 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYM 959
Db      894 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYM 953
Qy      960 IMVKCMWIDSECRPFRELIVSEFSMAADPORFVYIQNEDIGPASPLDSTFYRSLLEDDO 1019
Db      954 IMVKCMWIDSECRPFRELIVSEFSMAADPORFVYIQNEDIGPASPLDSTFYRSLLEDDO 1013
Qy      1020 MGDIYDAEEYLVPOGCFECPDPAAGAMVHRRSSSTRSGGDLTLGEPSEEPAPRS 1079
Db      1014 MGDIYDAEEYLVPOGCFECPDPAAGAMVHRRSSSTRSGGDLTLGEPSEEPAPRS 1073
Qy      1080 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTC 1133
Qy      1140 SPQPEYVNPQDVPRPQSPRESEPLPAARPAAGATLERAKTSLSPKNGVYKDVFAFGAVEN 1199
Db      1134 SPQPEYVNPQDVPRPQSPRESEPLPAARPAAGATLERAKTSLSPKNGVYKDVFAFGAVEN 1193
Qy      1200 PEYLTPQGGAAAPQHPPPAFSPAFNLYWDDPBERGAPSTFKGTPTAENPEYLGIDV 1259
Db      1194 PEYLTPQGGAAAPQHPPPAFSPAFNLYWDDPBERGAPSTFKGTPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

```

RESULT 11

AAE24067 standard; Protein; 1255 AA.

```

ID      AAE24067 standard; Protein; 1255 AA.
AC      AAE24067;
DT      23-SEP-2002 (first entry)
XX      Human Her-2 protein.
DE      Human Her-2 protein.
XX      Human, Her-2; epidermal growth factor receptor 2; infection; cancer;
KW      hyperproliferative disorder; propylaxia; inflammation; antisense;
XX      tumour; gene therapy; phosphorothioate backbone.
OS      Homo sapiens.
XX      PN
XX      MO200222636-A1.

```

```

PD      21-MAR-2002.
XX      PF
XX      12-SEP-2001; 2001MO-US28572.
XX      PR
XX      15-SEP-2000; 2000US-0663834.
XX      PA
XX      (ISIS-) ISIS PHARM INC.
XX      PI
XX      Bennett CF, Cowse LM;
XX      WPI; 2002-471192/50.
XX      DR
XX      N-PSDB; AAD38904.
XX      PT
XX      Novel antisense oligonucleotide which modulates the expression of Human
XX      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX      inflammation or to prevent infection in humans -
XX      Example 13; Page 95-107; 116pp; English.
XX      TS
XX      The invention relates to antisense compounds targeted to a nucleic
XX      acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX      that specifically hybridizes with and inhibits the expression of Her2.
XX      Antisense compounds of the invention are used for treating diseases or
XX      conditions associated with Her2 such as hyperproliferative disorders
XX      e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX      neural or cardiac cancer. They are also useful prophylactically e.g.
XX      to prevent or delay infection, inflammation and tumour formation. The
XX      invention is also used in gene therapy. The present sequence is human
XX      Her2 protein.
XX      SQ
XX      Sequence 1255 AA;
XX      Query Match 98.2%; Score 6730; DB 23; Length 1255;
XX      Best Local Similarity 98.4%; Pred. No. 0;
XX      Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2

```

```

Db      534  FLRGQECVEECRYLQGIIPREYVNAHCLPCHPECQPNQNSVTCFGEADQCAVACHAYXDP 593
Qy      600  PFCVACRPSGVKEDLSYMPIMKFPDEGACQCPICRICHSTCVLDLDKGPAPORASPLTS 659
Db      594  PFCVACRPSGVKEDLSYMPIMKFPDEGACQCPICRICHSTCVLDLDKGPAPORASPLTS 653
Qy      660  IVSAVVGILLVVLGVVFGILLIKRROOKIRKYMTRRLLOETELVEBELPSGAMPQAOQR 719
Db      654  IISAVVGILLVVLGVVFGILLIKRROOKIRKYMTRRLLOETELVEBELPSGAMPQAOQR 713
Qy      720  ILKTELIRKRVKVLGSGAFGTVMYKGIWIPDGENKIVAIKVLRNNSPANKELIDEAVY 779
Db      714  ILKTELIRKRVKVLGSGAFGTVMYKGIWIPDGENKIVAIKVLRNNSPANKELIDEAVY 773
Qy      780  MAGVSPYVSRLLIGITCTSTVQVLTQIMPYGCLLDHVRNCRGLSQDILNMQOIAJAKM 839
Db      774  MAGVSPYVSRLLIGITCTSTVQVLTQIMPYGCLLDHVRNCRGLSQDILNMQOIAJAKM 833
Qy      840  SYLEDVFLVHDLAARVNLVKSNNHVKITDFGLARLLDIDETETAHADGKVPDKMALES 899
Db      834  SYLEDVFLVHDLAARVNLVKSNNHVKITDFGLARLLDIDETETAHADGKVPDKMALES 893
Qy      900  ILRRFTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
Db      894  ILRRFTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
Qy      960  IMVKCMMIDSECRPRFRELVSERMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDOD 1019
Db      954  IMVKCMMIDSECRPRFRELVSERMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDOD 1013
Qy      1020  MGDVDAEEYLVPOQGFCEPDPAFGAGMWHHRSSSTRSGGDLTLGLESEEBEAPRS 1079
Db      1014  MGDVDAEEYLVPOQGFCEPDPAFGAGMWHHRSSSTRSGGDLTLGLESEEBEAPRS 1073
Qy      1080  PLAPSEGAGSDVFDGDLGMAANGLOSPLTHDPSLPQRYSEDPYVLPSETDGYVAPLTC 1139
Db      1074  PLAPSEGAGSDVFDGDLGMAANGLOSPLTHDPSLPQRYSEDPYVLPSETDGYVAPLTC 1133
Qy      1140  SPOEYVNOQDVNRQPPSPREGPLPAARPGATLEBPAKTLSPKNGVAVDVAFGAVEN 1199
Db      1134  SPOEYVNOQDVNRQPPSPREGPLPAARPGATLEBPAKTLSPKNGVAVDVAFGAVEN 1193
Qy      1200  PEYLTPOGGAPOHPHPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLTGLDV 1259
Db      1194  PEYLTPOGGAPOHPHPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLTGLDV 1253
Qy      1260  PV 1261
Db      1254  PV 1255

RESULT 12
AAE20479 standard; Protein; 1255 AA.
ID   AAE20479;
AC   AAE20479;
DT   01-JUL-2002 (first entry)
XX   Human Her-2/neu protein.
XX   Human Her-2/neu protein; immune response; gene therapy; breast cancer;
XX   human leukocyte antigen; HLA; vaccine; malignancy; cytosolic.
XX   Homo sapiens.
XX
XX   Key
XX   Region
XX   Location/Qualifiers
XX   1021..1030
XX   /note="Naturally processed HLA-B44-restricted epitope"
XX
XX   MO200214503-A2.

```

```

PD      21-FEB-2002.
XX
XX      14-AUG-2001; 2001WO-0641733.
PF
PR      14-AUG-2000; 2000US-225152P.
PR      28-SEP-2000; 2000US-226428P.
PR      21-FEB-2001; 2001US-270520P.
XX
XX      (CORI-) CORIXA CORP.
XX
XX      Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
XX      McNeill PD, Vedvick TS;
XX      WPI; 2002-280758/32.
DR      N-PSDB; AAD32743.
XX
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX      prevention and diagnosis of cancer, preferably breast cancer
XX
XX      Disclosure; Page 114-117; 129pp; English.
XX
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
XX
XX      Sequence 1255 AA;
XX
XX      Query Match 98.2%; Score 6730; DB 23; Length 1255;
XX      Best Local Similarity 98.4%; Pred. No. 0;
XX      Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy      1 MELDALCRWGLLALLPFGAASGVCTGDMKRLPASPETHLDMLRLHYGSCQVVGNTL 60
Db      1 MELDALCRWGLLALLPFGAASGVCTGDMKRLPASPETHLDMLRLHYGSCQVVGNTL 60
Qy      61 ELTYLPTNASTFLQDIOEVQGVVLLAHNOVQVPLQRLIRVGTQLFEDNALAVLDNG 120
Db      61 ELTYLPTNASTFLQDIOEVQGVVLLAHNOVQVPLQRLIRVGTQLFEDNALAVLDNG 120
Qy      121 DPLNNTPTVYVAGSPGGRLRELQSLTEILKGGVLIQNRNQLCYODTILMKDIFHNQNL 180
Db      121 DPLNNTPTVYVAGSPGGRLRELQSLTEILKGGVLIQNRNQLCYODTILMKDIFHNQNL 180
Qy      181 LTLIDNRSRACHPCSPCKSGKSWESSESDQSLTRYVCAAGCARCKPRLTDCCHQC 240
Db      181 LTLIDNRSRACHPCSPCKSGKSWESSESDQSLTRYVCAAGCARCKPRLTDCCHQC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICEHLGALVTYNTDTFESMPNPEGRTTFASCVTACP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICEHLGALVTYNTDTFESMPNPEGRTTFASCVTACP 300
Qy      301 YNYLSTDVSGCTLVCPHLNQEVTAEADGTORCEKSKPCARVCGYGLGEMNFTVSWLR-V 359
Db      301 YNYLSTDVSGCTLVCPHLNQEVTAEADGTORCEKSKPCARVCGYGLGEMNFTVSWLR-V 359
Qy      360 PKVSASHLEEFACCKIPQSLAFLEPSPFGDPAANTAPQEPQLOVFETLEITGYLTIS 419
Db      360 PKVSASHLEEFACCKIPQSLAFLEPSPFGDPAANTAPQEPQLOVFETLEITGYLTIS 419

```

```

Db      354  RAVTSANIQEFAGCKTIFGSLAFLPESFDGDPASNTAPLQPEQLQVETLEITGYLAYS 413
Qy      420  AMPDSLPLSLSPQULQVIRGRILNNGAYSLTLQQLGSMGLSLBELSGLLIHHNH 479
Db      414  AMPDSLPLSLSPQULQVIRGRILNNGAYSLTLQQLGSMGLSLBELSGLLIHHNH 473
Qy      480  LCFVHTVPMQDLFENPHQALHTANRPEDECEVSGGLACHOLCARGHCMGPGPTQVCNSQ 539
Db      474  LCFVHTVPMQDLFENPHQALHTANRPEDECEVSGGLACHOLCARGHCMGPGPTQVCNSQ 533
Qy      540  FLRGECEVECEKRVLOGLPREYVNAHCLPCHECECPONSVYTCGPPADCCVCAHAKCP 599
Db      534  FLRGECEVECEKRVLOGLPREYVNAHCLPCHECECPONSVYTCGPPADCCVCAHAKCP 593
Qy      600  PFCVARCPSGVKPDLSYMPWKPFDEEGACOPPCINCTHSCVDLDDKGCABEASAPLTS 659
Db      594  PFCVARCPSGVKPDLSYMPWKPFDEEGACOPPCINCTHSCVDLDDKGCABEASAPLTS 653
Qy      660  IVSAVVGILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQAQR 719
Db      654  IISAVVGILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQAQR 713
Qy      720  ILKETELRKRVVLGSGAFGYKGIWTPDGENYKIPVAIKVLSENTSPKAKKELDEAVY 779
Db      714  ILKETELRKRVVLGSGAFGYKGIWTPDGENYKIPVAIKVLSENTSPKAKKELDEAVY 773
Qy      780  MAGVSPVYSRLGLCLNTVQLVLTQMLPYGCLLDHVENRGRIGSODLLNMCQIADKM 839
Db      774  MAGVSPVYSRLGLCLNTVQLVLTQMLPYGCLLDHVENRGRIGSODLLNMCQIADKM 833
Qy      840  SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETETHAAGGVPIKMMALSS 899
Db      834  SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETETHAAGGVPIKMMALSS 893
Qy      900  ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIAREIPDLEKGERLPQPICTIDVYM 959
Db      894  ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIAREIPDLEKGERLPQPICTIDVYM 953
Qy      960  IMVKCMIMIDSECRPFRELVSEFSRMAADPQRFVITQNEDLGPAASPLDSTFFYSLLDD 1019
Db      954  IMVKCMIMIDSECRPFRELVSEFSRMAADPQRFVITQNEDLGPAASPLDSTFFYSLLDD 1013
Qy      1020  MGDLDVAEEYLVPOQGFPCPPAPGAGGMVHRHRSSTSGGDLTLGLEPSEEEAPRS 1079
Db      1014  MGDLDVAEEYLVPOQGFPCPPAPGAGGMVHRHRSSTSGGDLTLGLEPSEEEAPRS 1073
Qy      1080  PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPVPIPSETDGYVAPLTC 1139
Db      1074  PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPVPIPSETDGYVAPLTC 1133
Qy      1140  SPOPEYVNOQDVPRPSPREGPIPAARPAAGATLEBAKTLSPGKNGVVKVPAFGAVEN 1199
Db      1134  SPOPEYVNOQDVPRPSPREGPIPAARPAAGATLEBAKTLSPGKNGVVKVPAFGAVEN 1193
Qy      1200  PEYLTPQGAAPQHPPPAFSPAFDNLVYWDQDPBEGAPSTFKGTPTANPEYLGLDV 1259
Db      1194  PEYLTPQGAAPQHPPPAFSPAFDNLVYWDQDPBEGAPSTFKGTPTANPEYLGLDV 1253
Qy      1260  PY 1261
Db      1254  PY 1255

```

RESULT 13
AAMS1143
ID AAMS1143 standard; Protein; 1255 AA.

AC AAMS1143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

```

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KM tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
XX Homo sapiens.
OS
XX Key
XX Key Location/Qualifiers
XX FT 1..653
XX Domain /note= "extracellular domain"
XX FT 676..1255
XX Domain /note= "intracellular domain"
XX FT 990..1255
XX Domain /note= "phosphorylation domain"
XX
XX WO200212341-A2.
XX
XX 14-FEB-2002.
XX
XX 03-AUG-2001; 2001WO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX
XX (CORI-) CORIXA CORP.
XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Cheever MA, Gheysen D;
XX
XX WPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
XX or enhancing an immune response to the protein, has Her-2/neu
XX extracellular domain fused to Her-2/neu intracellular or
XX phosphorylation domain
XX
XX Claim 68; Fig 7; 14pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to EGF. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines
XX comprising the fusion proteins or nucleic acid molecules. In
XX preferred fusion proteins, the extracellular domain of a Her-2/neu
XX protein is fused to a Her-2/neu intracellular domain or
XX phosphorylation domain (or its DeltaCD fragment). An immune
XX response to Her-2/neu protein is elicited or enhanced by
XX administering the fusion protein in the form of a vaccine, or by
XX transfecting cells of an animal ex vivo with a nucleic acid
XX encoding the fusion protein, and delivering the transfected cells
XX to the animal. The fusion proteins, nucleic acids, and isolated
XX specific T-cells are useful for inhibiting the development of a
XX cancer, especially breast, ovarian, colon, lung or prostate cancer
XX in a patient. T cells that specifically react with a Her-2/neu
XX fusion protein can be used to remove tumour cells from a sample in
XX order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA:

```

Query Match 98.2%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2

Qy 1 MELALCRWGLLALLPPGAASIVCGTDMKRLPASPETHLDMRHLYOGGVQVGNL 60
Db 1 MELALCRWGLLALLPPGAASIVCGTDMKRLPASPETHLDMRHLYOGGVQVGNL 60

QY 61 ELYTPNASTSLFDIOIEVOGYVLIANNOVPLQRLRIVRGTOIFEDNYALAVLDNG 120
 DB 61 ELYTPNASTSLFDIOIEVOGYVLIANNOVPLQRLRIVRGTOIFEDNYALAVLDNG 120
 QY 121 DELNNTTPTVGASPGGLRELOLRSLTELKGGVLIQRPOLCYODTLMDIFHNNOLA 180
 DB 121 DELNNTTPTVGASPGGLRELOLRSLTELKGGVLIQRPOLCYODTLMDIFHNNOLA 180
 QY 181 LTLIDTNRSRACHPCSFWCKGSRGWSSESDCQSLTRTVCAAGGACRCKPLPTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSFWCKGSRGWSSESDCQSLTRTVCAAGGACRCKPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTNTDPEFMPBERGYFGASCVTACR 300
 DB 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTNTDPEFMPBERGYFGASCVTACR 300
 QY 301 YNYLSTDVGSCTLVCPLNHOEYTAEDGTQCEKSKPCARCYGLGMFNNFTVSMFLR-V 359
 DB 301 YNYLSTDVGSCTLVCPLNHOEYTAEDGTQCEKSKPCARCYGLGMFNNFTVSMFLR-V 359
 QY 360 PKVSASHLEEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPEOLQVETLEBITGYLYIS 419
 DB 360 PKVSASHLEEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPEOLQVETLEBITGYLYIS 419
 QY 420 AMPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLRSLRELGSGLALIHNTH 479
 DB 420 AMPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLRSLRELGSGLALIHNTH 479
 QY 480 LCFVHTVPMDOLEFNNPHQALHTANRPDEBCVGGGLACHOLCARGHGWPQTCVNCQ 539
 DB 480 LCFVHTVPMDOLEFNNPHQALHTANRPDEBCVGGGLACHOLCARGHGWPQTCVNCQ 539
 QY 540 FLRGQCEVCECRVQGLPREYVNAHCLPCHPEOPONGSVTCGEPADQCAVAKXP 599
 DB 540 FLRGQCEVCECRVQGLPREYVNAHCLPCHPEOPONGSVTCGEPADQCAVAKXP 599
 QY 600 PFCVACRCSGVKPDLSYMPIMKFPDEBGAQCPCHINCHSCVDLDKGCBAEGRASPLTS 659
 DB 600 PFCVACRCSGVKPDLSYMPIMKFPDEBGAQCPCHINCHSCVDLDKGCBAEGRASPLTS 659
 QY 660 TVSAVVGILLVVVLGVVFGLIKRQOKIRKTYRRLLOETELVEPLTBSGAMPNOQAKR 719
 DB 660 TVSAVVGILLVVVLGVVFGLIKRQOKIRKTYRRLLOETELVEPLTBSGAMPNOQAKR 719
 QY 720 ILKETELKXKVLGSGAGTVYKGIWIPDGENVKIPAIKVLRENTSPKXKEIIDEAYV 779
 DB 720 ILKETELKXKVLGSGAGTVYKGIWIPDGENVKIPAIKVLRENTSPKXKEIIDEAYV 779
 QY 760 MAGVSPYVSRLGLCLTSTVQLVTLQMPYGCLLDHVRENRGRIGSODLLNMCQIAKGM 839
 DB 760 MAGVSPYVSRLGLCLTSTVQLVTLQMPYGCLLDHVRENRGRIGSODLLNMCQIAKGM 839
 QY 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDPEGLARLDDIETEVHADGGVPIKMMALES 899
 DB 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDPEGLARLDDIETEVHADGGVPIKMMALES 899
 QY 900 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 959
 DB 900 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 959
 QY 960 IMVKCMWIDSCRFREFLVSFESRMARDPQRFVYIQNEDLGASPLDSTFFYSLLEDD 1019
 DB 960 IMVKCMWIDSCRFREFLVSFESRMARDPQRFVYIQNEDLGASPLDSTFFYSLLEDD 1019
 QY 1020 MGDLDVABEYLVPQGFCDPAPAGAGMYHHRSSSTRSGGDDTLIGLEPSEBAPRS 1079
 DB 1020 MGDLDVABEYLVPQGFCDPAPAGAGMYHHRSSSTRSGGDDTLIGLEPSEBAPRS 1079
 QY 1080 PLASEGAGSDVFEGDLMGAKGLQSLPTHDPSPFLQRYSEDPVLPSESTDGYVAPLTC 1139
 DB 1080 PLASEGAGSDVFEGDLMGAKGLQSLPTHDPSPFLQRYSEDPVLPSESTDGYVAPLTC 1139
 QY 1140 SPOEYVNVQPDVRFQPSPREGPIPARPAGATLERAKTILSPGNGVAVDVFAGAVEN 1199

DB 1134 SPOEYVNVQPDVRFQPSPREGPIPARPAGATLERAKTILSPGNGVAVDVFAGAVEN 1199
 QY 1200 PEYLTPQGAAPQHPFPAPFAPDNLYYMQDPEEGAPPSTFKGPTAENPEYLGLDV 125
 DB 1194 PEYLTPQGAAPQHPFPAPFAPDNLYYMQDPEEGAPPSTFKGPTAENPEYLGLDV 125
 QY 1260 PV 1261
 DB 1254 PV 1255
 RESULT 14
 AA077114
 ID AA077114 standard; Protein, 1255 AA.
 XX
 XX AA077114;
 DT 05-JUN-2002 (first entry)
 XX
 DE Human Her-2/neu polypeptide.
 XX
 KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
 KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 KW Hodgkin's lymphoma; T cell therapy.
 XX
 OS Homo sapiens.
 XX
 PN WC020213847-A2.
 XX
 PD 21-FEB-2002.
 PF 13-AUG-2001; 2001WO-US25408.
 PR 14-AUG-2000; 2000US-0638280.
 PR 28-SEP-2000; 2000US-0675904.
 PA (CORI-) CORIXA CORP.
 PI Gaiger A, Cheever MA, Hand-zimmermann S;
 XX WPI, 2002-280741/32.
 DR N-PSDB; ABK10730.
 XX
 PT Inhibiting haematological malignancy development by administering
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 PT polypeptide
 PS
 PS Disclosure; Page 71-74; 74pp; English.
 XX
 CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX
 SQ Sequence 1255 AA;
 Query March 98.2%; Score 6730; DB 23; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2

```

Db      61 ELTYLPTNASLSFLDIOEVQGVYLIAHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy      121 DELNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DELNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHQC 240
Db      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHQC 240
Qy      241 AAGCTGPKHSDCLACHENHSGICELHCPALVTTNTDTFESMPPEGRYTFGASCVTACP 300
Db      241 AAGCTGPKHSDCLACHENHSGICELHCPALVTTNTDTFESMPPEGRYTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPQLHNOEVTAEADGTORCEKSKPCARVCYGLGMEH-----LREV 353
Db      301 YNYLSTDVGSCTLVCPQLHNOEVTAEADGTORCEKSKPCARVCYGLGMEH-----LREV 353
Qy      360 PKVASHLEEFAGCKKIFGSLAFLESEFDGPASNTAPLOEQLOVETLEITGYLYIS 419
Db      360 PKVASHLEEFAGCKKIFGSLAFLESEFDGPASNTAPLOEQLOVETLEITGYLYIS 419
Qy      420 AMPDSLPLDSVFQNIQVTRGRIHNGAYSILTOGLIGISWLGRLSRLRELSGALIHNNTH 479
Db      420 AMPDSLPLDSVFQNIQVTRGRIHNGAYSILTOGLIGISWLGRLSRLRELSGALIHNNTH 479
Qy      479 LCFVHTVPMDQLFRNPQALHTANRPEDECVEGLACHQICAGHCGMPPTQCVNCSQ 539
Db      479 LCFVHTVPMDQLFRNPQALHTANRPEDECVEGLACHQICAGHCGMPPTQCVNCSQ 539
Qy      540 FLRGOECVEECVLOGLPREYVNAHCLPCHPECPQNGSTTCGPEPADQVCAAHAKDP 599
Db      540 FLRGOECVEECVLOGLPREYVNAHCLPCHPECPQNGSTTCGPEPADQVCAAHAKDP 599
Qy      599 PCVACRCPGKVPDLSTYMPIMKFPDEBAGACOPCINCHSCVDLDDKCCPRAGASPLTS 659
Db      599 PCVACRCPGKVPDLSTYMPIMKFPDEBAGACOPCINCHSCVDLDDKCCPRAGASPLTS 659
Qy      660 IYSAVVGILLVVLGVVFGILIKRQOKIRKYMRLIOETELVEPLTPSGAMNQOMR 719
Db      660 IYSAVVGILLVVLGVVFGILIKRQOKIRKYMRLIOETELVEPLTPSGAMNQOMR 719
Qy      720 ILKETELRKVKYLGSGAFSTYKGIWIPGENVKIPVAIKVIRENTSPKAKELIDAVY 779
Db      720 ILKETELRKVKYLGSGAFSTYKGIWIPGENVKIPVAIKVIRENTSPKAKELIDAVY 779
Qy      780 MAGVGSPIYSRLIGICLTSTVOLVTOLMPYGCILDHVENGRUGSODLLMCMQIAKGM 839
Db      780 MAGVGSPIYSRLIGICLTSTVOLVTOLMPYGCILDHVENGRUGSODLLMCMQIAKGM 839
Qy      840 SYLEDVRLVHRDLAARNVLVKSNNHVKITDPEGLALDIDETEVADGCKPIKMALES 899
Db      840 SYLEDVRLVHRDLAARNVLVKSNNHVKITDPEGLALDIDETEVADGCKPIKMALES 899
Qy      900 ILRRRFTHQSDVMSYGVTWELMTGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVM 959
Db      900 ILRRRFTHQSDVMSYGVTWELMTGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVM 959
Qy      960 IMYKCMWIDSECRPRRRELVSFESRMADPORFVVIQNEIDIGPASPDLSTYRSLDDDD 1019
Db      960 IMYKCMWIDSECRPRRRELVSFESRMADPORFVVIQNEIDIGPASPDLSTYRSLDDDD 1019
Qy      1020 MGDVDAEEYLVFOQGFPCPDAPAGAGVHHRHSSSTRSGGDLTLGLEPSEEAAPRS 1079
Db      1020 MGDVDAEEYLVFOQGFPCPDAPAGAGVHHRHSSSTRSGGDLTLGLEPSEEAAPRS 1079
Qy      1080 PLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSTLQSYSEDPYPLPSEFDGVAPLTC 1139
Db      1080 PLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSTLQSYSEDPYPLPSEFDGVAPLTC 1139
Qy      1140 SPOEYVNOQDVPKPPSPREGPLPAARPAAGATLERATLSPGKGVGVYDFAFGAVEN 1199
Db      1140 SPOEYVNOQDVPKPPSPREGPLPAARPAAGATLERATLSPGKGVGVYDFAFGAVEN 1199

```

```

Db      1134 SPOEYVNOQDVPKPPSPREGPLPAARPAAGATLERKTLSPKNGVYKDVAFGAVEN 115
Qy      1200 PEYLTPQGGAAPOHPPPAPFARFNDLYYMDQDPPERGAPSTFKGPTAEENPEYGLDV 125
Db      1194 PEYLTPQGGAAPOHPPPAPFARFNDLYYMDQDPPERGAPSTFKGPTAEENPEYGLDV 125
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KM Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93MO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston IL, Huston JS, Oppermann H, Ring DB;
DR WPI, 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083; AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 97.6%; Score 6687; DB 14; Length 1433;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 125; Conservative 10; Mismatches 9; Indels 8; Gaps 1;

Qy      1 MEALALCKWGLIALLPAGASTVCTGTDKLLPASPETHLMLRHLYQGCVVQGNL 60
Db      1 MEALALCKWGLIALLPAGASTVCTGTDKLLPASPETHLMLRHLYQGCVVQGNL 60
Qy      61 ELTYLPTNASLSFLDIOEVQGVYLIAHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLDIOEVQGVYLIAHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy      121 DELNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DELNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHQC 240
Db      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHQC 240

```

Search completed: July 22, 2003, 08:41:17
Job time : 43.1589 secs

```
QY 241 AAGCTGPKHSDCLACLFHNSGI CELHCPALVTYNTDTEESMNPBGRYTFGASCTYACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGI CELHCPALVTYNTDTEESMNPBGRYTFGASCTYACP 300
QY 301 YNYISTDVGSCTLVCPILHNOEVTAEOTQCEKCKSPCARVCYGLMFNNFTVSFMR-V 359
Db 301 YNYISTDVGSCTLVCPILHNOEVTAEOTQCEKCKSPCARVCYGLMFNNFTVSFMR-V 359
QY 360 PKYSASHLEEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPEOLQVFETLEETGYLYIS 419
Db 360 PKYSASHLEEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPEOLQVFETLEETGYLYIS 419
QY 354 RAYTSANIGEFAGCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLQVFETLEETGYLYIS 413
Db 354 RAYTSANIGEFAGCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLQVFETLEETGYLYIS 413
QY 420 AMPDSLPLDSVPQNLQYTRGRILHNGAYSITLQGLGISMGLRSLELGSGLALIHNTH 479
Db 420 AMPDSLPLDSVPQNLQYTRGRILHNGAYSITLQGLGISMGLRSLELGSGLALIHNTH 479
QY 414 AMPDSLPLDSVPQNLQYTRGRILHNGAYSITLQGLGISMGLRSLELGSGLALIHNTH 473
Db 414 AMPDSLPLDSVPQNLQYTRGRILHNGAYSITLQGLGISMGLRSLELGSGLALIHNTH 473
QY 480 LCFVHTVPMDOLEFNPQALLHTANRPEDECVGEGLLACHQLCARHGWGPGPTCCVNCQ 539
Db 474 LSFVHTVPMDOLEFNPQALLHTANRPEDECVGEGLLACHQLCARHGWGPGPTCCVNCQ 533
QY 540 FLRGQECVEECRVLOGLPREYVNAHCLPCHBECQPNQSVTCFGEADQVACAHYXP 599
Db 534 FLRGQECVEECRVLOGLPREYVNAHCLPCHBECQPNQSVTCFGEADQVACAHYXP 593
QY 600 PCVCARCPGKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTS 659
Db 594 PCVCARCPGKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTS 653
QY 660 IVSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 719
Db 654 IISAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 713
QY 720 ILKETELRKVKVJLGSAGFTVYKGIWIPDGENYKIPVAILRENTSPKANKEILDEAYV 779
Db 714 ILKETELRKVKVJLGSAGFTVYKGIWIPDGENYKIPVAILRENTSPKANKEILDEAYV 773
QY 780 MAGVSPYVSRLLGICLTSTVQCLVTOLEMPYGCLLDHRNRRGLSGODLLNMCQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVQCLVTOLEMPYGCLLDHRNRRGLSGODLLNMCQIAKGM 833
QY 840 SYLEDVRLVHRLAARNLVKSPNHVKITDFGLARLLIDETEHADGGKVPKXMALES 899
Db 834 SYLEDVRLVHRLAARNLVKSPNHVKITDFGLARLLIDETEHADGGKVPKXMALES 893
QY 900 ILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYM 959
Db 894 ILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVCMWIDSECRPRRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDD 1019
Db 954 IMVCMWIDSECRPRRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDD 1013
QY 1020 MGDLYDAEEYLVPOQGFCDPAPAGAGWVHHRSSSTRSGGDLTLGLEPSEEDAPRS 1079
Db 1014 MGDLYDAEEYLVPOQGFCDPAPAGAGWVHHRSSSTRSGGDLTLGLEPSEEDAPRS 1073
QY 1080 PLAPSEGASDVFDGDLGMAKGLQSLPTHDPSPLOKXSEDPVLEPSETGYVAPLTC 1139
Db 1074 PLAPSEGASDVFDGDLGMAKGLQSLPTHDPSPLOKXSEDPVLEPSETGYVAPLTC 1133
QY 1140 SPOPEYVNOQDVAPOPSPREGPLPARPAGATLERPKTISPGKNGVVDVAFGAVEN 1199
Db 1134 SPOPEYVNOQDVAPOPSPREGPLPARPAGATLERPKTISPGKNGVVDVAFGAVEN 1193
QY 1200 PEYLTPGGGAAPQHPHPAPSPAFNDLYWDDODPPERGAPESTFKGTPAENREYLGLDV 1259
Db 1194 PEYLTPGGGAAPQHPHPAPSPAFNDLYWDDODPPERGAPESTFKGTPAENREYLGLDV 1253
QY 1260 PV 1261
Db 1254 PV 1255
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds

(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MELAALCRMGILLALLPPGHA.....TRKGTPTAENPYLGIDVVP 1261

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: piri: *
2: piri: *
3: piri: *
4: piri: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Query Length	DB ID	Description
1	6655	97.1	1255	1 A24571	protein-tyrosine k
2	5853	85.4	1260	1 TVRINU	protein-tyrosine k
3	5846.5	85.3	1254	1 T48161	p-185 precursor
4	3101	45.3	1210	1 GQHUE	epidermal growth f
5	3075	44.9	1223	1 A53183	epidermal growth f
6	2934.5	42.8	1203	1 TVCHLV	epidermal growth f
7	2934.5	42.8	1308	1 A47253	epidermal growth f
8	2639	38.5	1166	1 S06142	protein-tyrosine k
9	2382.5	34.8	1342	2 A36232	kinase-related tra
10	2284.5	33.5	1339	2 UC487	kinase-related tra
11	1766.5	25.8	699	1 TVFVLV	epidermal growth f
12	1703	24.9	604	1 TVYOH	protein-tyrosine k
13	1647	24.0	544	2 S35745	protein-tyrosine k
14	1640	23.9	545	2 S00727	protein-tyrosine k
15	1623	23.7	540	2 B44776	kinase-related tra
16	1621	23.5	540	1 TVFVEB	protein-tyrosine k
17	1613.5	23.5	1330	1 GQFE	epidermal growth f
18	1467	21.4	644	2 A3625	epidermal growth f
19	1285	18.8	1333	2 E88257	epidermal growth f
20	1285	18.8	1374	2 S70712	protein let-23 (im
21	1202	17.5	1369	2 S70713	protein-tyrosine k
22	1179	17.2	1369	2 A45558	epidermal growth f
23	1089	15.9	527	2 A42032	epidermal growth f
24	958.5	14.0	843	2 A27131	epidermal growth f
25	806.5	11.8	846	2 S13807	epidermal growth f
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	733	10.7	1363	2 T43200	insulin-like growt
28	698	10.2	1382	1 INHUR	insulin receptor p
29	696	10.2	1607	2 T43212	insulin-like growt

30	691	10.1	1383	2 A36080	insulin receptor p
31	690.5	10.1	1372	2 A34157	insulin receptor p
32	680	9.9	1477	2 T18534	protein-tyrosine k
33	667.5	9.7	1300	2 A36502	insulin receptor-r
34	658	9.6	1268	2 B36502	insulin receptor-r
35	636	9.3	1367	1 IGHORI	insulin-like growt
36	626	9.1	1350	2 T30346	insulin receptor -
37	624	9.1	1371	2 A33637	insulin-like growt
38	619.5	9.0	2148	1 A56081	insulin receptor
39	615	9.0	2101	2 S57245	insulin receptor
40	596.5	8.7	987	2 A54092	protein-tyrosine k
41	588	8.6	1114	1 S05582	protein-tyrosine k
42	586	8.6	1091	2 S33596	protein-tyrosine k
43	585	8.5	976	2 A36355	protein-tyrosine k
44	583.5	8.5	952	2 I50612	protein-tyrosine k
45	581.5	8.5	977	2 S49004	tyrosine kinase Mp

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming pr
C/Species: Homo sapiens (man)
C/Date: 25-Oct-1987 #sequence, revlusion 06-Dec-1996 #ext, change 11-Jun-1999

C/Accession: A24571; A25491; A4188; B44188; I59509; I57622
R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; S

Nature 319, 230-234, 1986
A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epiderma

A/Reference number: A24571; M01D:86118653; PMID:3003577
A/Accession: A24571

A/Molecule type: mRNA
A/Residues: 1-1255 <YAM>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6487-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB
A/Reference number: A25491; M01D:86016729; PMID:2959567

A/Accession: A25491
A/Molecule type: DNA

A/Residues: 737-1031 <SEM>
A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Consens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGarath, J.;
Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shar
A/Reference number: A44188; M01D:86070181; PMID:2999974

A/Accession: A44188
A/Molecule type: DNA

A/Residues: 740-910 <COU1>
A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188
A/Molecule type: mRNA

A/Residues: 1-517, 'PALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 228, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcin
A/Reference number: I59509; M01D:85212597; PMID:2992089

A/Accession: I59509
A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 832-909 <REX>

A/Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R/Fal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987
A/Title: Human HBR2 (neu) promoter: evidence for multiple mechanisms for trans

A/Reference number: I57622; M01D:87286898; PMID:3039551
A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA

A/Residues: 1-191 <PAL>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetic: A:Gene: GDB:ERBB2; NGL; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function: A:Description: catalyzes the phosphorylation of a peptide tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:22-653/Domain: extracellular #status predicted <EXT>
 F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F:195-605/Domain: EGF receptor extracellular domain repeat <EB2>
 F:554-675/Domain: transmembrane #status predicted <TM>
 F:576-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68-124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:553/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.1%; Score 6655; DB 1; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 8,66-266;
 Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2;

```

QY      1 MELALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLMDLHLYOGCCVVOGNTL 60
DB      1 MELALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLMDLHLYOGCCVVOGNTL 60

QY      61 ELTYLPINASLSFLDIDIEVOGVYLIANQVROYLQRLRTVRGTQLFEDNYALAVDNG 120
DB      61 ELTYLPINASLSFLDIDIEVOGVYLIANQVROYLQRLRTVRGTQLFEDNYALAVDNG 120

QY      121 DPLNNTTPVTGASPGQLRELQRLSTELIKGVLIQRNPQLCYODTILMKDIFPKNNQLA 180
DB      121 DPLNNTTPVTGASPGQLRELQRLSTELIKGVLIQRNPQLCYODTILMKDIFPKNNQLA 180

QY      181 LTLIDTNSRACHPSKPCSKGSRGSGESSEDCOSLTRVACGAGCARCKGPPPTDCCHEQC 240
DB      181 LTLIDTNSRACHPSKPCSKGSRGSGESSEDCOSLTRVACGAGCARCKGPPPTDCCHEQC 240

QY      241 AAGCTGPKHSDCLACLAHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
DB      241 AAGCTGPKHSDCLACLAHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300

QY      301 YNYLSTDVGSCTLVNQLNOEVTADGQREKSKRCARVCYGLMFPNNFTVSWIR-V 359
DB      301 YNYLSTDVGSCTLVNQLNOEVTADGQREKSKRCARVCYGLMFPNNFTVSWIR-V 359

QY      360 PYVASHLSEFAGCKKIFGSLAFLPESFGDPASNTAPQVYKANSKFIQTTELTYGLYIS 419
DB      360 PYVASHLSEFAGCKKIFGSLAFLPESFGDPASNTAPQVYKANSKFIQTTELTYGLYIS 419

QY      420 AMPDSLPLDSVFQNTQVTRGRILHNGAVSLTGLGISMGLSLRLSGGLALIHNNTH 479
DB      420 AMPDSLPLDSVFQNTQVTRGRILHNGAVSLTGLGISMGLSLRLSGGLALIHNNTH 479

QY      474 LCFVHTVPMQDLFRPHQALLHTANRPEDEVGSLGCHQICARHGHWGSPQVCVNCQ 533
DB      474 LCFVHTVPMQDLFRPHQALLHTANRPEDEVGSLGCHQICARHGHWGSPQVCVNCQ 533

QY      540 FLRGOECVEBCRVYQGLPREYVNAHCLPCHPRECQPNQSVTCGPEPADQVCAAHYKDP 599
DB      540 FLRGOECVEBCRVYQGLPREYVNAHCLPCHPRECQPNQSVTCGPEPADQVCAAHYKDP 599

QY      534 FLRGOECVEBCRVYQGLPREYVNAHCLPCHPRECQPNQSVTCGPEPADQVCAAHYKDP 593
DB      534 FLRGOECVEBCRVYQGLPREYVNAHCLPCHPRECQPNQSVTCGPEPADQVCAAHYKDP 593

QY      600 PFCVAPCEPGVYKPDLSYMPIMKPFDEBGAQPCPINCTHSCVLDLDDGCAEGRASPLTS 659
DB      600 PFCVAPCEPGVYKPDLSYMPIMKPFDEBGAQPCPINCTHSCVLDLDDGCAEGRASPLTS 659

```

```

DB      594 PFCVAPCEPGVYKPDLSYMPIMKPFDEBGAQPCPINCTHSCVLDLDDGCAEGRASPLTS 655
QY      660 IYSANVGLLVVVLGVVGLIILKROCKIRKRYTKRLLQETELVEPLTPSGAMPNQOMR 715
DB      654 IISAVVGLLVVVLGVVGLIILKROCKIRKRYTKRLLQETELVEPLTPSGAMPNQOMR 715
QY      720 ILKETELRKVKLGSGAGTYKGIWIPDGENVKI PAIKYLRNTPSPKANKELLDPAVY 775
DB      714 ILKETELRKVKLGSGAGTYKGIWIPDGENVKI PAIKYLRNTPSPKANKELLDPAVY 775
QY      780 MAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHYENKRGSLGODLLNMCQIAKGM 835
DB      774 MAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHYENKRGSLGODLLNMCQIAKGM 835
QY      840 SYLEVRVLRVHDLAARNVLYKSPHNVKITDGLARLDDIETEVHADGCVPIKMMLES 895
DB      834 SYLEVRVLRVHDLAARNVLYKSPHNVKITDGLARLDDIETEVHADGCVPIKMMLES 895
QY      900 ILRRRFTQSDVWSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 955
DB      894 ILRRRFTQSDVWSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 955
QY      960 IMVKCMITDSECRPRFRLVSESSRMARDPQRFVYIONEDLGPASPLDSTFYSLLED 101
DB      954 IMVKCMITDSECRPRFRLVSESSRMARDPQRFVYIONEDLGPASPLDSTFYSLLED 101
QY      1020 MODLVDAEYLVPOQGFPCPDPAAGAGMWHRRSSSTSGGDLTLGLSPSEEAAPS 107
DB      1014 MODLVDAEYLVPOQGFPCPDPAAGAGMWHRRSSSTSGGDLTLGLSPSEEAAPS 107
QY      1080 PLAPBEGAGSVFPGDLCMGAKQLQSLPTHPSPLQRYSEDPVLPUSSTDGYVAPLTC 113
DB      1074 PLAPBEGAGSVFPGDLCMGAKQLQSLPTHPSPLQRYSEDPVLPUSSTDGYVAPLTC 113
QY      1140 SPOPEYVQCPVRQPPSPRPGPLPARPGATLTERKTLSPGKNGVYKQVAFGAVEN 115
DB      1134 SPOPEYVQCPVRQPPSPRPGPLPARPGATLTERKTLSPGKNGVYKQVAFGAVEN 115
QY      1200 PEYLPQGAAPQHPPPAFSPAFDNLVYMDODPPERGAPSTIFKGTPTAENEYGLADV 125
DB      1194 PEYLPQGAAPQHPPPAFSPAFDNLVYMDODPPERGAPSTIFKGTPTAENEYGLADV 125
QY      1260 PY 1261
DB      1254 PY 1255

```

RESULT 2
 TYRNT
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Barzmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A>Title: The neu oncogene encodes an epidermal growth factor receptor-related
 A:Reference number: A24562; MUID:86118662; PMID:3945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.
 Carcinogenesis 12, 1975-1978, 1991
 A>Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain re
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:920525293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics: A:Gene: neu

C: Superfamily: epidermal growth factor receptor; protein kinase homology
C: Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-13/Domain: signal sequence #status predicted <SIG>
F:120-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:731.191.263.555.576.634/Binding site: carboxylate (asn) (covalent) #status predicted
F:691/Binding site: phosphate (thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:682.1227.1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.4%; Score 5853; DB 1; Length 1260;
Best Local Similarity 85.7%; Pred. No. 6.5e-233;

Matches 1083; Conservative 58; Mismatches 113; Indels 10; Gaps 4;

```
1  MELAALCRWGLLALLPFGAASVCTGTDKMLRLPASPETHLDMRLHYOGCQVQGNL 60
4  MELAAMCRWGLLALLPFGIAGTGTGTDKMLRLPASPETHLDMRLHYOGCQVQGNL 63
61  ELTYLPNASTLSFLODIOEVQGVYLIAHNOVROPRLRIYVGTQLFEDNYALAVLDNG 120
64  ELTYLPANASTLSFLODIOEVQGVYMLIAHNOVKRVPRLRIYVGTQLFEDKVALAVLDNR 123
121  DPLNNTTPVT-GASBGGIRELQLRSLTEILKGVLIQGNPOLCYQDTILMKDIFHKNNQL 179
124  DPQNVASTPGRRPEGRELQLRSLTEILKGVLIQGNPOLCYQDMVLMKDVFRKNQL 183
180  ATLITDTRSPACHPCSPCKGSRCKWBSSEDCOSLTRVACAGGACRCKPLPTDCHEQ 239
184  APVDIDTRSPACHPCACPKCKDNHCKWBSSEDCOILTGITCTSGACRCKRLPTDCHEQ 243
240  CAAGCTGPKHSDCLACHFNHSG:CELCPCALVYNTDTESMPNPEGRYTFGASCVTAC 299
244  CAAGCTGPKHSDCLACHFNHSG:CELCPCALVYNTDTESMPNPEGRYTFGASCVTAC 303
300  PYNVLSTDVGSCTLVCPILHNGEVTAEQGTORCEKSCPCARVCYGLGMFNFTVSEFMLEV 359
304  PYNVLSTEVGSCTLVCPNNOEVTAEQGTORCEKSCPCARVCYGLGMFNFTVSEFMLEV 356
360  PK-VSASHLEEFACKCKIFGSLAFSPFODDASNTAPQYIANSKSFITLTGILYI 418
357  ARLTSDVQEFDDCKKIFGSLAFSPFODDASNTAPQYIANSKSFITLTGILYI 416
419  SAMPDSLPLDSVFNLOVIRGRILHNGAVSLTQGLISWLGRLSRLREGSLALIHNT 478
417  SAMPDSLPLDSVFNLOVIRGRILHNGAVSLTQGLISWLGRLSRLREGSLALIHNT 476
479  HLCVHTVPMQDLFRNPHQALLTANRPEDB-CVBSGLACHQLCARGHCWGPGPTQCVNC 537
477  HLCVHTVPMQDLFRNPHQALLHSGNRPEBDLCVSSGLVCNSLCARGHCWGPGPTQCVNC 536
538  SCFLRGOECVECEVQLQGLPREYVNAHCLPCHEQCPQNGSVTCGPREADQCVAAHAK 597
537  SHFLRGOECVECEVQWGLPREYVSDKRCCLPCHEQCPQNGSVTCGSEBDDQCAAHAK 596
598  DPFPCVACPSGVKPDLSYMPWKFPDEGACQPCPINTCHSCVDDLDDKCPAEPQASPL 657
597  DSSCVACPSGVKPDLSYMPWKFPDEGACQPCPINTCHSCVDDLDDKCPAEPQASPL 656
658  TSISAVAVGILLVVLGVFGILLIKRQOKIRKYTRKRLLOSTELVEPLTPSAMNQO 717
657  TSISAVAVGILLVVLGVFGILLIKRQOKIRKYTRKRLLOSTELVEPLTPSAMNQO 716
718  MRILKETELRKVYLGSGAGTVYKGIWIPDGENVKIPVAIKYLRANTSPKANKELIDBA 777
717  MRILKETELRKVYLGSGAGTVYKGIWIPDGENVKIPVAIKYLRANTSPKANKELIDBA 776
778  YVMAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHRENNRGSLGSDLLNMCQIAK 837
777  YVMAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHRENNRGSLGSDLLNMCQIAK 836
838  GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDTEGLARLLDIDETEVHADGKVPKIMMAL 897
```

```
837  GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDTEGLARLLDIDETEVHADGKVPKIMMAL 896
898  ESILRRRTHOSDVWSYGVYWE:MTFGAKPYDGI:PAE:IPDLLEKERLP:POPI:CTIDV 957
897  ESILRRRTHOSDVWSYGVYWE:MTFGAKPYDGI:PAE:IPDLLEKERLP:POPI:CTIDV 956
958  YMIWKCMIDSECRPPRELVEFSFSSMAADPQRFVYIQNEDLGAPSLDSTYRSLLED 101
957  YMIWKCMIDSECRPPRELVEFSFSSMAADPQRFVYIQNEDLGAPSLDSTYRSLLED 101
1018  DDMGDLVAEEYLPQGGFPCCDDAPAGCMVHHRHSSSTRSGGDLTLGLPSEEGAP 107
1017  DDMGDLVAEEYLPQGGFPCCDDAPAGCMVHHRHSSSTRSGGDLTLGLPSEEGAP 107
1078  RSPILAPSEGASDVFDGDLGKAAGLQSL:PTHDPSPLQRYSEDPYVLP:PSSETDGYAPL 113
1077  RSPILAPSEGASDVFDGDLGKAAGLQSL:PTHDPSPLQRYSEDPYVLP:PSSETDGYAPL 113
1138  TCSPPPEYVNPQDVRFQPSRREGPLPAAPPAKTLERAKTLSPGXGVYKDVFAFGAV 115
1137  ACSPPPEYVNPQDVRFQPSRREGPLPAAPPAKTLERAKTLSPGXGVYKDVFAFGAV 115
1198  ENPEYLPQGGAAQDPHPAPFSPAFNLLYWMQDPPERGAPSTFGPTAENPEYLG 125
1197  ENPEYLPQGGAAQDPHPAPFSPAFNLLYWMQDPPERGAPSTFGPTAENPEYLG 125
1258  DVPV 1261
1257  DVPV 1260
RESULT 3
148161
P:185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.;
Gene 140, 261-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; PMID:94193007; PMID:7908275
A:Accession: 148161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BA03801.1; PID:9747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
Query Match 85.3%; Score 5846.5; DB 2; Length 1254;
Best Local Similarity 85.3%; Pred. No. 1.2e-232;
Matches 1077; Conservative 69; Mismatches 107; Indels 9; Gaps 3
```

```
1  MELAALCRWGLLALLPFGAASVCTGTDKMLRLPASPETHLDMRLHYOGCQVQGNL 60
1  MELAAMCRWGLLALLPFGIAGTGTGTDKMLRLPASPETHLDMRLHYOGCQVQGNL 60
61  ELTYLPNASTLSFLODIOEVQGVYLIAHNOVROPRLRIYVGTQLFEDNYALAVLDNG 120
61  ELTYLPANASTLSFLODIOEVQGVYMLIAHNOVKRVPRLRIYVGTQLFEDKVALAVLDNR 120
121  DPLNNTTPVTGASBGGIRELQLRSLTEILKGVLIQGNPOLCYQDTILMKDIFHKNNQL 180
121  DPLNNTTPVTGASBGGIRELQLRSLTEILKGVLIQGNPOLCYQDTILMKDIFHKNNQL 180
181  ATLITDTRSPACHPCSPCKGSRCKWBSSEDCOSLTRVACAGGACRCKPLPTDCHEQ 240
181  ATLITDTRSPACHPCACPKCKDNHCKWBSSEDCOILTGITCTSGACRCKRLPTDCHEQ 240
181  PVDIDTRSPACHPCACPKCKDNHCKWBSSEDCOILTGITCTSGACRCKRLPTDCHEQ 240
```

QY 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVTYNTDFFESMPNDEGRYTFGASCYACP 300
 DB 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVTYNTDFFESMPNDEGRYTFGASCYACP 300
 QY 301 VNYLSTDVSGCTLVCPPLNHNQETVADTGQRCCKSPCARVCYGLGPNFTYFWLRVP 360
 DB 301 VNYLSTDVSGCTLVCPPLNHNQETVADTGQRCCKSPCARVCYGLGPNFTYFWLRVP 360
 QY 361 K-VASASHEEPAGCKKIFGSLAFLEPESFGDPASNTAPQYIKANSKFIQITELTGYLIS 419
 DB 361 K-VASASHEEPAGCKKIFGSLAFLEPESFGDPASNTAPQYIKANSKFIQITELTGYLIS 419
 QY 420 AMPDSLPLDLSVFQNTQVIRGRILLNNGAVSLTQGLISWLGSLRLRELGSGLALIHNTH 479
 DB 420 AMPDSLPLDLSVFQNTQVIRGRILLNNGAVSLTQGLISWLGSLRLRELGSGLALIHNTH 479
 QY 480 LCFVHTVPWDOLEFRPHQALLHTARPEDECVBESLACHQICAGHCHMGPEPTQVNCVCSQ 539
 DB 480 LCFVHTVPWDOLEFRPHQALLHTARPEDECVBESLACHQICAGHCHMGPEPTQVNCVCSQ 539
 QY 540 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQCVACAHYKDP 599
 DB 540 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQCVACAHYKDP 599
 QY 600 PCVVARCBGVKVPDLSTYMPIMKFPDEBEGACQPCPINCCHSCVDDDKGCPABQASPLTS 659
 DB 600 PCVVARCBGVKVPDLSTYMPIMKFPDEBEGACQPCPINCCHSCVDDDKGCPABQASPLTS 659
 QY 660 IYSAVVGILLVVLVVGILIKRROQKIRKYTWRLIQETELVEPIPTSGAMNQAKMR 713
 DB 660 IYSAVVGILLVVLVVGILIKRROQKIRKYTWRLIQETELVEPIPTSGAMNQAKMR 713
 QY 720 ILKETELRKVVLGSGAGFTVYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 779
 DB 720 ILKETELRKVVLGSGAGFTVYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 779
 QY 780 MAGVGSPIVSRLLGICLTSTVQVLTQMLPYCCLLDHRENGRGRISODLLMWCQIAKGM 839
 DB 780 MAGVGSPIVSRLLGICLTSTVQVLTQMLPYCCLLDHRENGRGRISODLLMWCQIAKGM 839
 QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGLARLLDIDETEVHADGKVPIMKMALES 899
 DB 840 SYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGLARLLDIDETEVHADGKVPIMKMALES 899
 QY 900 ILRRRFTHOSDVWVSGVYVWELMTGAPYGCITAREIPDLLENGERLPQPPICITIDVYM 959
 DB 900 ILRRRFTHOSDVWVSGVYVWELMTGAPYGCITAREIPDLLENGERLPQPPICITIDVYM 959
 QY 960 IMVKKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1019
 DB 960 IMVKKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1019
 QY 1020 MGDLYDABEYLVPOGFCFCPPDAPAGAMVTHRRSSSTRSGGDLTLGLEPSEEAPRS 1079
 DB 1020 MGDLYDABEYLVPOGFCFCPPDAPAGAMVTHRRSSSTRSGGDLTLGLEPSEEAPRS 1079
 QY 1074 PLAPSEGAAGSDVFBELGMAATKGPQGISLPPLOQISDPLPLPPTETDGVVAPLAC 1133
 DB 1074 PLAPSEGAAGSDVFBELGMAATKGPQGISLPPLOQISDPLPLPPTETDGVVAPLAC 1133
 QY 1140 SPQPEYVNOQDVRRPQSPSPREGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAVEN 1199
 DB 1140 SPQPEYVNOQDVRRPQSPSPREGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAVEN 1199
 QY 1200 PEYLTPOGGAAPQPPPPAFSPADNLYYWDQDPEEGAPSTKSGPTLENREYGLDQV 1259
 DB 1200 PEYLTPOGGAAPQPPPPAFSPADNLYYWDQDPEEGAPSTKSGPTLENREYGLDQV 1259
 QY 1253 PV 1254
 DB 1253 PV 1254

RESULT 4
 GENE
 epidermal growth factor receptor precursor - human
 N/contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C/species: Homo sapiens (man)
 C/date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
 C/accession: A00641; A25772; S30024; A38672; A00642; A23062; A05281; J
 R/Julich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.;
 rg, P.H.
 Nature 309, 418-425, 1984
 A/title: Human epidermal growth factor receptor cDNA sequence and aberrant ex
 A/reference number: A00641; MUID:84219729; PMID:6328312
 A/accession: A00641
 A/molecule type: mRNA
 A/residues: 1-1210 <URL>
 A/cross-references: EMBL:X00588; NID:g31113; PIDN:CAA5240.1; PID:g757924
 A/Note: the authors translated the codon AAG for residue 540 as Asn
 R/Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
 A/title: Characterization and sequence of the promoter region of the human ep
 A/reference number: A25772; MUID:85270438; PMID:2991899
 A/accession: A25772
 A/status: translation not shown
 A/molecule type: DNA
 A/residues: 1-29 <ISH>
 A/cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
 R/Haley, J.D.; Whitte, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield
 Oncogene Res. 1, 375-396, 1987
 A/title: The human EGF receptor gene: structure of the 110 kb locus and ident
 A/reference number: S30024; MUID:86217333; PMID:3329716
 A/accession: S30024
 A/molecule type: DNA
 A/residues: 1-29 <HA2>
 A/cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
 R/Haley, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A/title: Contributory effects of de Novo transcription and premature transcrip
 A/reference number: A38672; MUID:91107677; PMID:1998448
 A/accession: A38672
 A/molecule type: DNA
 A/residues: 1-29 <HA1>
 A/cross-references: GB:M38425; NID:g181977; PIDN:AAA61771.1; PID:g553271
 A/Experimental source: carcinoma cell line A431.7
 R/Xu, Y.; Ishii, S.; Clark, A.D.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe,
 Nature 309, 806-810, 1984
 A/title: Human epidermal growth factor receptor cDNA is homologous to a variet
 A/reference number: A00642; MUID:84245835; PMID:6330563
 A/accession: A00642
 A/molecule type: mRNA
 A/residues: 150-187, 'KSVYQNV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'R',
 '798-799, 'TD', 802-811, 'R', 813-942 <XUY>
 A/Experimental source: A431 human carcinoma cells, which have large numbers of
 R/Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.;
 Science 224, 843-848, 1984
 A/title: Expression cloning of human EGF receptor complementary DNA: gene ampl
 A/reference number: A43615; MUID:84196372; PMID:6326261
 A/accession: A43615
 A/molecule type: mRNA
 A/residues: 713-964 <LIN>
 A/Experimental source: epidermoid carcinoma cell line A431
 R/Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malle
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A/reference number: A23062; MUID:85046483; PMID:6093780
 A/accession: A23062
 A/molecule type: mRNA
 A/residues: 1028-1210 <SIM>
 R/Weber, W.; Gill, G.N.; Speiser, J.
 Science 224, 294-297, 1984
 A/reference number: A05281; MUID:84172183; PMID:6524343
 A/accession: A05281
 A/molecule type: Protein
 A/residues: 25-30, 'S', 32-51, 454-467 <WEB>

R,Russo, M.W.; Lukas, T.J.; Cohen, S.; Stamos, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A>Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A/Reference number: A60143; MUID:85182650; PMID:2985580
 A/Accession: A60143
 A/Molecule type: Protein
 A/Residues: 740-744,'X',746-747 <RUS>
 R/Mrozowski, B.; Nosiq, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A/Reference number: A38023; MUID:84191554; PMID:6325948
 A/Accession: A38023
 A/Contents: annotation: receptor activity
 A/Notes: The EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R/Chen, W.S.; Lazar, C.S.; Lund, K.A.; Weis, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain 1
 A/Reference number: A33331; MUID:90003233; PMID:2790960
 A/Accession: A33331
 A/Contents: annotation: internalization signal
 C/Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C/Genetics:
 A/Gene: GDB:EGFR
 A/Cross-references: GDB:120610; OMIM:131550
 A/Map position: 7p12.3-7p12.1
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine; signal sequence; status predicted <SIG>
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <MAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:666-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128,175,352,413,444,528,603/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:745/Active site: Lys #status experimental

Query Match 45.3%; Score 3101; DB 1; Length 1210;
 Best Local Similarity 49.3%; Pred. No. 4e-120;
 Matches 626; Conservative 175; Mismatches 358; Indels 112; Gaps 23;

```

QY 11 LLLALLPPGAA--STOVCTGTDMKLLPASPETHLDMRLHYOGQVQVQNELTYLPTN 68
DB 14 LLAALCPASALEBKVKVCGSTSKLQLGTFFDHFLSLQRMNNEVVLGNIEITYVRN 73
QY 69 ASUSFLQDIOEVGVYLIANOVROYPLQRLKIVGTQLFEDNYALAVLDNGDPLNITTP 128
DB 74 YDLSFLKTIQEVAVGVYLIANTVERPLENLQIRGNMYENSVALAVLSNYD----- 126
QY 129 VTGASPGGLRELOSLSTELKGVLIQENPOLCYODITLMDIFHKNNQLALTLDTNR 188
DB 127 ---ANKTGLKELPMKRLQELHGAVRFSNNPALCNVESIQMDIVSSDPLSNMWDFONH 183
QY 189 SRACHPCSPWCKSRGWGSSSDQSLRTVCAGCA-RCKGPLRLTDCHEGCAAGCTGP 247
DB 184 LGSCKCKDPSCPWGSQWAGGEENCKLTKIIQAQCSGCRGKSPDSCCHNQCAAGCTGP 243
QY 248 KHSDDLACLHFNHSGICELHCPALVYNTDFESNPNBGRTRFASCVTACPNVYSTD 307
DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNTFYQVMDVNPGRKSFATVKKCPRRYVYTD 303
QY 308 VGSCTLVCPLEHNOEVAEDGTQRCCKSPKCARVCGYGLGMFNFTVSWLBPVKVASHL 367
DB 304 HGSQVACGADSYEM-EEQGVKCKCKCEBPCKVCGNGIGI-GEFDSL-----SINATNI 356
QY 368 EEFAGCKTIFGSLAFIPESFPDDPASNTAPQIKANSKFTIGTLETGLYISAMPDCLPD 427
DB 357 KGFNCTISIGDHLIPVAFRGDSFTHPPDLPQEDLTKVKEITGGLLIQAMPENRTD 416
QY 428 LSVFQMLQVIRGLIHNGAVSLTQGLIGISLGLSLRELISGLALIHNNTHLCFVHTVP 487

```

```

DB 417 LHAENLEIRGRTKOHGFSLAVSNLNTSLGRSLKEISDGVIIISGNKNLCYANTIN 476
QY 488 MDOLFERNPHOALHTANRPDECEVSGGLACHOLCAAGHCWGPGPTQVCVNSQFLRQECV 547
DB 477 WKTLFGISGQKTIKINRGENSKATQVCHALCSPEGCGWGBEPDVCVCRVNSRGREV 536
QY 548 EECRVYGLPREVYVNAHCLPCHPECOPONGSVTCEGPADOCVACAHYKDPFCVACRP 607
DB 537 DKCKLLEGREFREVENSECTIQCHPECLPQAMNITCTGPRDNCICCAHYIDGPHCVKTCR 596
QY 608 SGVPRDLSTVMPKFPDEBACQPCPGINTCTHSCVDLDDGCAEQRASLTIVSNVVG- 666
DB 597 AGVWGENNTL-VKTVADAGVCHLCPNCTCYCTGCTGCGLECPNPKLP--SIARGWGA 653
QY 667 --LLVYLVGVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQMRILKET 724
DB 654 LLLLVVALGIG---LFMRRLHVRKTRRLRLQRELVLEPLTPSGEARNQALLIKET 710
QY 725 ELARKVYLSGAGFYTKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDBAVYVAGV 784
DB 711 EFKKIKVLSGAGFYTKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDBAVYVAGV 770
QY 785 SPVYSRLIGTCLSTVQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV 844
DB 771 NPHVCRLLIGTCLSTVQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQ 830
QY 845 VRLVHDLAARNVLYKSPHNVKITFGLARLLDIDETEHADGKVPKIMWLESILRR 904
DB 831 RRLVHDLAARNVLYKSPHNVKITFGLARLLDIDETEHADGKVPKIMWLESILRR 890
QY 905 FTHQSDVSGVTVWELMTFGAKRPNYGIPIAREIPLLEKGERLPPRPCTIDVYIMWKC 964
DB 891 YTHQSDVSGVTVWELMTFGAKRPNYGIPIAREIPLLEKGERLPPRPCTIDVYIMWKC 950
QY 965 WMTDSECRPFRELVSEFARMARDPQRFVYIQ-NEDLGPASLDSTFYSLEDDMDGL 102
DB 951 WMTDASRKFRRLIIEFGRKARDPQRYVIVIGDGRMLPSTDSNFRALMDEEDMDV 1014
QY 1024 VDAEYLVYPOQGFCCDPAFAGAGVNHRRSSSTRSGGDLTLGLSEEBEAPRPLAP 108
DB 1011 VDAEYLVYPOQGFCCDPAFAGAGVNHRRSSSTRSGGDLTLGLSEEBEAPRPLAP 1034
QY 1084 SEGAGSDVDFDGLGMAAGKLSLPHDPSLPQRYSEDPVYLPSET--DGVVALTCLSP 114
DB 1037 SLSATSN--NSTVACIDRNGLOSCEPKEDSLQRYSSDPTGALTDSIDDTFL-----P 1086
QY 1142 QPEYVNPDPVRQPPSPREGRPLPAARPAATLERAKTILSPGKGVYKDVFAFGAVENPE 1201
DB 1089 VPEYVING-SVPRKPADSVQNPVYHNPQINP-----APSRDPHYQD--PHSTAVGPE 1137
QY 1202 YL-TPQGGAAAPQHPPRPAFPAFNDLYWDQ-----DP-----PERGAPSTFK 1244
DB 1138 YLNTVQ-----PTCVNSTFDSPAHMAQKSHQISLNDPDDYQDFFPEAKENGIFK 1186
QY 1245 GTPTEAPPEYL 1255
DB 1189 GS-TAENAYL 1198

```

RESULT 5
 A53183
 Epidermal growth factor receptor precursor - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C/Accession: A53183; A43818; S24942; A28941; S43525; I49643
 R/Letter(s): N.C.; Phililips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkin
 Genes Dev. 8, 399-413, 1994
 A>Title: The mouse waved-2 phenotype results from a point mutation in the EGF r
 A/Reference number: A53183; MUID:94170986; PMID:8125255
 A/Accession: A53183
 A/Molecule type: mRNA
 A/Residues: 1-1210 <LUE>
 A/Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, U.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A.Reference number: A43818; MUID:91232866; PMID:2050916
A.Accession: A43818
A.Molecule type: mRNA
A.Residues: 1-714 <AVI>
A.Cross-references: GB:X59698
R:Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A.Reference number: S24942
A.Accession: S24942
A.Molecule type: mRNA
A.Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A.Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A>Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A.Reference number: A28941; MUID:80330814; PMID:3138233
A.Accession: A28941
A.Molecule type: protein
A.Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A>Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A.Reference number: S45325
A.Accession: S45325
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-971, 'K', 973-1210 <VER>
A.Cross-references: EMBL:X78987; NID:G488830; PIDN:CA55587.1; PID:G488831
R:Patia, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A>Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A.Reference number: I49643; MUID:93126380; PMID:7678348
A.Accession: I49643
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 12-20, 22-132 <RES>
A.Cross-references: GB:L06864; NID:G193001; PIDN:AAA5029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoproc
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-677/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:693/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.98; Score 3075; DB 2; Length 1210;
Best Local Similarity 49.28; Pred. No. 4, 7e-119;
Matches 630; Conservative 169; Mismatches 361; Indels 120; Gaps 26;

Qy	11	LLALLPPGAA--STOYCTGTDMKLRLPASPEHLDMLRHLVGGCCVQGNILETLPLPN	68
Db	14	LLTLLCAAAGLLEKKKCCGGSNRLTGLTFEDHFLSLQRMNNCCVVGNIETLVQGN	73
Qy	69	ASISFLQDIOEVQGVILIAHNOYRQVFLRLRIVRGTOLEFEDNVALVLDNDPPLNNTTP	128
Db	74	YDLSFLKTIQEVAGYVILAIINVERIPLENIQLIRGALYENTYALATLSN-----	124
Qy	129	VTGASPGGLRELQRLSTELLKGVILIQNRPOUCYODTLIMKI-----PKKNQLALITI	184
Db	125	YGTNRGLRELPRKNOETLIGAVFSSNPILCNMDTIQMDITVONVSKNSMDL---	180
Qy	185	DYNSRACHPCSPWCKSCRCMGSSSEDCQSLRTVYAGGCA-RCKGPLPTDCHEGCAAG	243
Db	181	QSHSSCPKCDPFCPCPGSCMGGEENCOGLTKIICAGQCSHRCRGRSPSDCHNCGAAG	239

Qy	244	CTGPKHSDCLALCFHNSGICELHCPALVTYNTDTFESMPNPGRTFGASCTYAPVNY	303
Db	240	CTGRBEDCLVCCQFQDEATCKTQCPRLMLYNTTYQMVONPBGKYSFATCKKCPRY	299
Qy	304	LSTDVGSCTLVCPLNHOEVTAEADGTORCEKSCPKCARVCGYGLM--FNNFTVSWLRVPK	361
Db	300	VYTDHGSQVACGPDYEV-EEDGIRKCKKCDGPKRCVNGIGIGFCK-TLS-----	350
Qy	362	VSASHLEEPGCKKIFQSLAFLEESPDGPASTAPQYKANSKFGITELTGLYISGM	421
Db	351	INATINHFYKCYAISGDLILFVAKFGSFTPLPDLRELEILKTVEITGFLILQAM	410
Qy	422	PDSLPDLSVFNQLOVIRGLIHNGAVSLTLOGIGIMLGRSLREKSGGALIHNTHTLC	481
Db	411	PDNNTDLHAFENLEIRGRTKQHGQSLAVGNTISLGRSLKEISDGDVILSGNRNIC	470
Qy	482	FVHTVPMQDLPNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFL	541
Db	471	YANTINMKKLFQTPNQTKIMNNRAEKDCAVHNVNCPSSSEGCWGPPEPDCVSCONVS	530
Qy	542	RGQECVEGRVLQGLPREYVNAHCLPCHPEOQPOKGSVTCGEPADOCVACHYKDPF	601
Db	531	RGKCEVCKNILEGPREFVENSEICQCHPECLPQAMNITCTGRGPDNCTICAHYIDGH	590
Qy	602	CVARCPGVKPDLSYPIWKFPEDEGACQPCPINCTHSCVDLDDKCCPAEORASPLTSTV	661
Db	591	CVKTCFPAIGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCEVWBSGPKIPSLA	649
Qy	662	SAVIGILLVYVVGAVGI-LIKRQOKIRKTYMRRLQETELVEPLTBSGAMPNOQMKI	720
Db	650	TGIVGGLLFTIV-VAGIGLGMRRRIIVKRTLRLLQRELEVEPLTBSGEAPNQAHRI	708
Qy	721	LKETELKRVKVLGSAFGVYVKIMIPDENVYKIPAIKYLENTSPKANKIILDAAYM	780
Db	709	LKEREFKIXIVLSGARFVYKLMIPBEKXKIPAIKELNATSPKANKIILDAAYM	768
Qy	781	AGVSPVYSRLGICLTSTVQVLTQMLPFGCCLDHYENRGRIGSQDILNMCQIAKGS	840
Db	769	ASVDNPFVCHRLGICLTSTVQVLTQMLPFGCCLDHYENRGRIGSQDILNMCQIAKGN	828
Qy	841	YLDEVRVNRDLAARVNVKSPHNVITTFGLARLIDIDETVHADGGVVPKMMALSTI	900
Db	829	YLDERLVNHDLAARVNVKTPPHVKITTFGLAKLGABEKYHAGGVVPKMMALSTI	888
Qy	901	LRRRFTHQSDVMSYGVYVLMFTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVMI	960
Db	889	LHRIYTHQSDVMSYGVYVLMFTFGSKPYDGIIPASDISILEKGERLPQPICTIDVMI	948
Qy	961	MYCKMTIDSCRRREFELVSEFRMRDQRFVITQ-NEDLGAPSLDSTFRSLLEDQ	101
Db	949	MYCKMTIDASRKFRELLIEFSQMRDQRFVITQGDERRMLPSPFTDSNFYRALMDEED	100
Qy	1020	MGDLVDAEYLVFQQGFPCDPAPFAGGAGVHNRHRSSTRSGGDLTLLESESEAPRS	107
Db	1009	MEDVDADEVLTIQQGF-----NSPST-----SRT	103
Qy	1080	PLAPSEAGSDVDGDLGMAAGLOSLETHDPSPLQRYSEDPVPLPSET--DGYVAPL	113
Db	1035	PLLSLSLSTSN---NSTVACINRNSCRVKDPAFLQRYSSDPTGAVTEDNIDDAFL---	108
Qy	1138	TCSFQPEYVQPDVRRPQSPREGRPLPAPRPAAGATLERAKTISPGNGVAVKVPFAGV	119
Db	1088	---FVPEYVQ--SVPRPAPGQVQNPYHNPQLHP-----APGRDLHYON--PNSNAV	113
Qy	1198	ENPEYL-TPOGGAAPQPHPPAPFAFDNLVYWDQ-----DP-----PERGAP	124
Db	1134	GNPEYLYNTAQ-----PTCLSSGFNSPALMIGSHQMSLNDPDYQODFPFKETFN	118
Qy	1241	STFKGPTAEPEYLGDPV	1260
Db	1185	GIRKG-PTAENAYLVAVP	1203

```

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
M:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C/Accession: A27720; #sequence_revision 05-May-1995 #text_change 04-Feb-2000
R:Lat, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ulrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85282822; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:61-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:65-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #
F:167/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.6%; Score 3057.5; DB 1; Length 1223;
Best Local Similarity 48.0%; Pred. No. 2,5e-118;
Matches 625; Conservative 175; Mismatches 352; Indels 151; Gaps 26;
QY 8 RWGLLLALLPPGAA-----STOVCTGTDMKRLRLPASPTHLDMLRHLYQGQGVVQGNLE 61
DB 13 RGAVALVLLLLGVALCSAVEEKVCCGTTNNKLTQLGHVEDHPTSLQRMYNCEVLLSNLE 72
QY 62 LTYLPTNASLSPLQDIEGVGYVLIANQVROYPLQRLRIYVGTQFFENVALATLNDG 121
DB 73 IITYEHRDLPLKTIQEVAGYVLIANVVDYPLLENQIRKNVLYDVSFLAALSNH 132
QY 122 PLNNTPTVTCASPGGIRELQRLSLTELKGVLIQNNPOLCYQDTLLMKDIPHKNNQAL 181
DB 133 -NMKTQ-----GLRELPMKRLSEILNGVXISNNPKLCMMMDLYLNMDDIDTSRK-PL 182
QY 182 TLID-TNRSRAHPCSPMKGRSCWGESSEDCSLRTYCAAGCA-RCGAPLPTCCHEQ 239
DB 183 TVDPASNLSCKPCGHNCCTEDHCWAGEQNCQTLTKVICAOQCSGRGKVPSSCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDFESMPNDEGRYTFGASCYTAC 299
DB 243 CAAGCTGPRESDCLACGKFRDADATCKDCEPLVLYPTTYQMDVNPBGYSFGALCYAEC 302
QY 300 PNYLSTDVSSCLVCPRLHNOETADGRCCKGKPCARVCYGLGMNFTVSFMLRV 359
DB 303 PHNYVVDHGSVCVRSQNTDTEYV-EENGVKCKCKCGGLSKVONGG-----IGELKGI 355
QY 360 PKYSASHLEEFAGCKKIFGSLAPLPSPFGDPPASNTAPCYIKANSKFGITELTYGLVIS 419

```

```

DB 356 LSIATNIDSFKNCTKINGDVSLIPVAFGLDAPFKTLPLDPKKLDVFRVTKVEISGLILIO 415
QY 420 AMPQSLDLSVFOQLQVIRGRILHNGAYSTLLOGLISWLGRLSELGSLALIHNNH 479
DB 416 AMPNADLDLAFELLETIRKTRQHQGYSALVNLKIQSLGSLSELISGDIALMKKN 475
QY 480 LCFVHTVPMDQLFBNPQALHTANPREDCEVGEGLAQQLCARHCWGPPTQVCNCSG 539
DB 476 LCVADTNMWSLFTQSGQKRIQNNKNDCTADRHVCPDLGSDVGCWGGPCHCSGCF 535
QY 540 FLRQCEVBEQVLCIIPREYVNAHCLPCHPCQYQNG---SVTFGPBADQVCACAH 596
DB 536 FSRQKECVKQCNILQGPREFEBSKCLPCHSECLQVNSTAVYNTQSGGPPDHCKMCAH 595
QY 597 KDPFCVACRPSGKPLSLYMPILPDEDEGACOPRINCSTHSCVLDLGGCAEGRASP 656
DB 596 IDGHCHKACRPAVGLGENDL-VKKAADANAVQLCHPNCRCRCKPGLGCF--NSGK 651
QY 657 LTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKYMRLLETELVEPLTPSGAMPNQ 715
DB 652 TPSIAGVVGGLLCLVVGGLGILYLR-R-HVRSKTLRRLQERELVEPLTPSGEAPNQ 710
QY 716 AQNRILKETELRKVYLSGAFGTVYKGIWIRGEMVYKIPVAIKULENTSPKANEILD 775
DB 711 AHRILKETEFKVKVLSGAFGTVYKGLMIRGEXVKIPVAIKULENTSPKANEILD 770
QY 776 EAYVMAVGSPPVSRLLGLCLSTVQLVTLQMLMYGCLLPHVENRGLSODLLAMCMQI 835
DB 771 EAYVMAVSDPHVCRLGLIGCLSTVQLVTLQMLMYGCLLDYIRHKNKISQVYLMNCVQI 830
QY 836 AKGMSYLEDRLVHRDLAANVLYKSPNNHYKITDPSLALLDIDETRYHADGGKVPITKM 895
DB 831 AKGMYLLEERLVRDLAANVLYKTPHYKITDPSLALLDIDEXEYHAEGKVPITKM 890
QY 896 ALBSILRRRPTHQSDVSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICTI 955
DB 891 ALBSILRRITHQSDVSVGVTVWELMTFGSKPYDGIIPAREISLVLEKGERLPPICTI 950
QY 956 DVTIMVYKCMIDSECRPRELIVSEFSMARDPQRFVVIQ-NEDLPASPDLSTFYRSL 101
DB 951 DVTIMVYKCMIDADSRPRELIVSEFSMARDPQRFVVIQ-NEDEPASPDLSTFYRSL 101
QY 1015 LEDDDMDLDVAREEYVLPQGFPCPPRACAGAGMHHRRSSSTSGSGDILLGLEPSEE 107
DB 1011 MEEEDMEDVDVADAEYVLPQGF-----NSBST----- 103
QY 1075 EAPRSPV-----APSEGAGSDVDDGLGMAAKGLQLPTHPSPLOQYSEDPVLPSE 112
DB 1039 --SRPLLSLSLSTSNNSATNCID-----RNGGHHVREDSPFORISSPTGNFLEE 108
QY 1130 T--DGYVAPLTCSPOEYVNPQVPRPQSPRBPGLPARPAGATLERAKTILSPGKGV 118
DB 1089 SIDDGFL-----PABEYVNO-LMKKES-----TAMQONQIY 111
QY 1188 KDVF-----AFGAVNPEYLLPQGGAPRPPAPFAFNTLYYMD- 123
DB 1120 NMSLRAISKLPMDSRYKSHANDNPEYL-----NTQSLPAKTVESSPYWIOS 117
QY 1232 -----DPE-----RGAPSTFKGTPTAENPEYGLDVP 1260
DB 1172 GNRQINLDNDYQDELPRNTRKNGILLKVPAAENPEYLRVAAP 1214

```

```

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C/Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Fe
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the e

```

A:Reference number: A47253; MUID:93189574; PMID:8383326
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1308 <PI0>
 A:Cross-references: GB:107868; NID:9337359; PIDN:AAB59446.1; PID:9337360
 A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase homology <KIN>
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.8%; Score 2934.5; DB 2; Length 1308;
 Best Local Similarity 44.7%; Pred. No. 2.9e-113;
 Matches 605; Conservative 188; Mismatches 380; Indels 179; Gaps 29;

QY 9 KGLLLALPFGAA-----STQVCTGIDMKLRIPASEETHLMDLRHLHYGCGVQVGNLELY 64
 DB 8 WWSVSLVAAGTVQPSDSQSVCAQENTLSSLSLDEQCYRALRKYENCCEVMGNLEITS 67
 QY 65 LPTNASLFLDIOQVGVLLAHNVQVPLQRLRIYRGTLFEDNYALAVLNDGDFLN 124
 DB 68 IEHNDLSFLRSVREVTGVVALNPFYLENLRIIRGKLYEDRYALALIFLNTRKDG 127
 QY 125 NTPVTVGASPGGLRELQRLSTEILKGVLLIQRPOLCYDPTIIMKDIPIKNNQALVTLI 184
 DB 128 NF-----GLOELGLKNTLEILNGGVYDQMKFLCYADTIHMQIVANPMSNLTIV 178
 QY 185 DTRRSRACHPGSPMKGSRGSESEDDQSLTRVVCAGGC-ARGKGLPTDCCHEQCAAG 243
 DB 179 STNGSSGGRCHKSTG-RCMGPTENHCQTLRIYCAEQDCGRGYGVYDCCHEQCAAG 237
 QY 244 CTGPKHSDCLALAHFNHSGICELHCPALVTYNTDTPESMPNDEGRYTFGASCVTACPYNY 303
 DB 238 CSGPKDTCFACMNPNDGACVTCQPTFYVYNPTTFQLEHNFNKAYTYGAFVCKKCPHNF 297
 QY 304 LSTDVSGCTIVCPHNOEVTADGTQREKSKPCARCYGLGFENNFTVSMRLRPKYS 363
 DB 298 V-VSSSCVRACPSKKEV-BENGKMKKPCCTDIPCACDGI-----TGSIMSQTVYD 349
 QY 364 ASHLEEFAGCKKIFGSLAFLEPESFDGPASNTAQYKANSKFIETELTGYLYISAMP 423
 DB 350 SSNIDKFINCTKINONLFLVTGHHGDPYNAIEADEKLVNFTVETIGFLINQSWP 409
 QY 424 SLPLDLSVQNLQVTRGRLIHGANSLSLQIGISLGLRSIRELSGLALIHNTHLCPV 483
 DB 410 NMTDFSVSNLVTIGRVLVYSGLSLILIKQOGITSLOFOSLKEISAGNIYITDNSNLCTY 469
 QY 484 HTVPMDQLFRNPHOALHTANRPEDECVGEGLACHOLCARGHGMGPPPTQCVNCSQFLG 543
 DB 470 HTINMTLFTIINGRIYRDNKRAKENTCAEGMVCNHLCSDDCKGPPDCLSCRRFSRG 529
 QY 544 QECVEECRVLOGLPREYVNAHCLPCHPECOP-ONGSVTCFGEADQVCAAHYKDPPEC 602
 DB 530 RICESCNLYGGEFEFENGSIQVECDPQCEKXEDGLTCHGROPNDCTKCSHFKGPMC 589
 QY 603 VARCPGVKPLSLVMPIMKPFDEGACOPGINTHSCVULDKG-----PAE 651
 DB 590 VEKCPDGLQGANSF--IFKYADPRECHPCHPNTQGCNGPTSHDCIYFPMTHSTLPH 647
 QY 652 QRASPLTSIVAVY-GILLVVVLGVFGILIKRQCKIRKYRRLLOETELVEPLTPSG 710
 DB 648 AR-TPL--IAAGVIGGLFILIYVGLTFAVYVRSIK-KKRALRFL-ELEVEPLTPSG 702
 QY 711 AMPNOQWRLIKETELKRVKYLGSAGACTYKGIWIDGNNVKIPVAILKRLNTSPKAN 770
 DB 703 TAPNOQRLIKETELKRVKYLGSAGACTYKGIWIDGNNVKIPVAILKRLNTSPKAN 762
 QY 771 KEIDEAAYVAGVSPYVSRLLGICLTSTVOLVTLMPYGCCLDHYRENGRGLSGDOLLN 830
 DB 763 VEFMDEALIMASMDHPHLVRLGLVCSPITQVLTQMLPHCLLEVEHKNINIGSGLLN 822
 QY 831 WCMQIAKMSYLEVRLVHNDLAARNVLVSPNHVKITDGLARLLDIDETEVHADGKV 890

DB 823 WCVQIAKMSYLEERLVRDLAARNVLVSPNHVKITDGLARLLDIDETEVHADGKM 88;
 QY 891 PIKMALESILRRRPTHOSDVMSYGVTWELMTGAPYDGPDIPIREIPDLLEKERLPOR 950
 DB 883 PIKMALESILRRRPTHOSDVMSYGVTWELMTGAPYDGPDIPIREIPDLLEKERLPOR 94;
 QY 951 PICTIDVYMWKCMKIDSECPREPRFELVSEFSMAWDPORFVYIQNED-LGPASPLDST 100
 DB 943 PICTIDVYMWKCMKIDSECPREPRFELVSEFSMAWDPORFVYIQNED-LGPASPLDST 100
 QY 1010 FYRSLLEDMDGDVDAEYVLPVQGFPCPPAPAGAGMHHRRSSSTSGGDLTLGL 106
 DB 1003 FQNLIDEDLEDMDDAEYVLPVQGFPCPPAPAGAGMHHRRSSSTSGGDLTLGL 106
 QY 1070 EPESEAPRS-----PLAP-SEGASDVFPDGLCMGA 110
 DB 1051 SFPAYTSMGNGFYVRDQGFPAAGVSVFAPRTSTIPAPVYQATAEIFDDSCNGT 111
 QY 1101 AKGLQSLPTHQPSPLQVSEDPYPLPS-----ETDGYVAPLTGSPQRYNQDPVR 115
 DB 1111 LRKVPAPHVQDSSSTQYSADPTVFAERSPPGELDEGYMTPEYRDKPKQELNPNVE-- 116
 QY 1154 QPPSPREGLPAPAPAGATLERAKTILSPKNGVYQVFAFGAVENPEYLTPOGGAAPOR 121
 DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNEFYHNASNG----- 115
 QY 1214 HPPPA-----FSPAFNLVYWDOPPERGA- 123
 DB 1195 -PPKADEYVNEPLVNTFANTLQKAEYLNKNNILSMPEKAKAFDNDPNHSLPFRSTL 125
 QY 1239 -PSTFKGTPT-----AENPEYL 1255
 DB 1254 QHPDLOEYSTRKYEVKQNGRIRPVAENPEYL 1285

RESULT 8
 506142
 protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
 N.Alternate names: epidermal growth factor receptor homolog; kinase-related tr
 C.Species: Xiphophorus maculatus (southern platyfish)
 C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 18-Feb-2000
 C.Accession: S06142; S13809
 R.Witbrodt, U. Adam, D. Maltzschek, B. Maenele, W. Raulf, F. Telling, A
 Nature 341, 415-421, 1989
 A.Title: Novel putative receptor tyrosine kinase encoded by the melanoma-induc
 A.Reference number: S06142; MUID:50015140; PMID:27977166
 A.Accession: S06142
 A:Molecule type: DNA
 A:Residues: 1-1166 <MIT>
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291
 R.Adam, D. Maenele, W. Scharf, M.
 Oncogene 6, 73-80, 1991
 A.Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in
 A.Reference number: S13807; MUID:9115882; PMID:1846957
 A.Accession: S13809
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 821-1025 'N', 1027-1098 'A', 1100-1166 <ADA>
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285
 A:Genetics:
 A:Gene: mrk
 A:Map position: Y
 A:Insertions: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane pro
 F:1-25/Domain: signal sequence #status predicted <Sig>
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted
 F:716-972/Domain: protein kinase homology <KIN>
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 38.5%; Score 2639; DB 1; Length 1166;
 Best Local Similarity 44.5%; Pred. No. 3.4e-101;

Matches 568; Conservative 173; Mismatches 384; Indels 152; Gaps 32;	
QY 4 AALCRMGILLALPPGAASST-----OVCGTDMKLRLPASPEHLDMLRHLYOGQGVQGN 59	QY 1074 EEARPSPLAPSGAGSDVFDGDLGMGAAGLQSLPTHPDSPLRYSDDPRV-P-LPSETDG 11:
DB 8 AALLQ--LLLVLSIRCCSTDPDRKVCQGTSNQMTM--LDNHTLKNKKMISGGNVLEN 62	DB 1020 -----EPCLPRTGH-----PVRENSTTLRNISDPTONALEKLDG 10:
QY 60 LELTYLPTNASISFLQDIQEVQGVYLLAHNOVROVPLQRLRIVRGTLFEDNVALAVLDN 119	QY 1133 YVAPLTCSPQPEEYVNPDPVRPP-----PSPRE-----GPLP-AARPGATLERAKTL 11:
DB 63 LEITYQENODLSFLQSIQEVQGVYLLAHNEVSTIPLVNLRLRIGQVLYEGNFTLVMSN 122	DB 1055 H-----EYVNPQSESTSSRLSDIYNPNYEDLTDGMCQPVSLSSQEAETNFSREYL 11:
QY 120 GDPINNTPTVTGASGGRLBELQRLSTELILKGVLQNLQNPCLCYQDTILMKDITHKNQL 119	QY 1180 SPQKNQVQKDVAFAGGAENPEYLLTPQGAAPQPHPRPASPARDNLVYNDQDPPEGCAP 12:
DB 123 YQK-NPSSP--DVYQVGLKQQLQSLNLTILSGYKASHNPLLCVETITNMWDIVDKTSNP 179	DB 1105 NTNQNSL--PLVSSGSDDDPY---QAG-----YQAF-----LPQTGL 11:
QY 180 ALTLIDYNSRACHPCSPMCKSGRCMGSESSDCSLRTFVQAGC-ARCKGPLPTDCHE 238	QY 1240 PSTFKGTPTAENPEYLQ 1256
DB 180 TNNLIPIHAFERCCQCHGCVNCGMARGPHCKFTLLCAEQCNRCRQPKRIDCNE 239	DB 1140 TQNGMFLPAENLEYLG 1156
QY 239 QCAAGCTGPKHSDCLACHFNHSGICEIHCPLVTTNTDTFESPNPEGRYTFGACVTA 258	RESULT 9
DB 240 HCAGGCTGPRATDCLACRDFNDGCKQTCPPKIIDVSHQVVDNPNIKYTFGACVKE 259	A36223
QY 299 CPYNYLSTPVGSCCTLYCPHLNQEVTAEDGTQCEKSCAPACVCGYGMFN-NFTVFWL 357	kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
DB 300 CPENYVTE-GACVASCAGMLEVD-ENGKRSCKPCQGVCKVCDGIGIGLSNTIA--- 354	C.Species: Homo sapiens (man)
QY 358 RVPKVSASHLEEFACCKIFQSLAFPSFDGP---ASNTAPQYIKANSKFIGITELTG 414	C.Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
DB 355 ----VNSTNIRSFNSTCTINGDIILNRSFEGDHYKIGTMDPEHL---WNLTIVKEITG 407	C.Accession: A36223; 159164
QY 415 YLISAMPDSDLDLSPONLOVIRGIRLIHNGAS-LTLQGGISMLRLSRISREGSLAL 473	R.Kraus, M.H.; Issing, W.; Miki, T.; Pospescu, N.C.; Aaronson, S.A.
DB 408 YLVIMMBENMTSLSVFQWLEIRKRTTFSGFSFVAVQVHQLMQLRSIKESASANVI 467	Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
QY 474 IHNHTLCEVHTVPMDOLFNRPHQALLHTANRPDEVCESGLACHQUCASGHCMGCPPTQ 533	A.Title: Isolation and characterization of ERBB3, a third member of the ERBB/
DB 468 LKNTQLKRYANTINMRLLFRSEDOIEYDART-----ENQTCNNCSDEGCM-PEPTM 519	A.Reference number: A36223; MUID:90083234; PMID:2687875
QY 534 CYNCSQFLRGECEVEECVQLQGLPREYVNAHCLPCHEPCQPNQSGSYTCGPEADQCVAC 593	A.Accession: A36223
DB 520 CVSCILHVRGRGVASCNLLQGEPEACVDRCVQCQOECLVQDLSITCYGFPGRANSKS 579	A.Status: preliminary
QY 594 AHYKDPFCVACPSGVNFDLSYMPIMKFPDEBACQPCPINCHTSCVDDDDKQCPAEQR 653	A.Molecule type: mRNA
DB 580 AHFDGPOCIRCRPHGILGDDGL-IMKYADKMQCQPCQCHNCQSGSGSLGCRD-1 637	A.Residues: 1-1342 <KRA>
QY 654 ASPLTISAVVGLILVVLGVVGEILIKRPQOKIRKYTMRELLIQETELVEPLTPSGAMP 713	A.Cross-References: GB:M29366
DB 638 VSHSLAVGLVSGLLITVIALTLVLLRRRRIR-KRRTIRCLIQEKELVEPLTPSGAP 696	R.Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.;
QY 714 NOAQRILKETELRKVKVLGSGAFGVYKGIWPDGENVKIPVAKYLRNTSPKANKER 773	Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
DB 697 NOAFIRILKETEFKDRVLGSGAGTGYKGLMNDGENIRIPVAKYLRNTSKNQEV 756	A.Title: Molecular cloning and expression of another epidermal growth factor
QY 774 LDEAVYVAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYENENGRLSGSDLNMCW 833	A.Reference number: 159164; MUID:90311312; PMID:2164210
DB 757 LDEAVYVAVSVHPVCRLLGICLTSAVQVLTQMLPYGCLLDYVHQHERICGQWLLNMCV 816	A.Accession: 159164
QY 834 QIAKMSYLEDVRLVHPRLAARVLYSPNHVKITTDGLARLDDIDETEAAGKVPK 893	A.Status: preliminary
DB 817 QIAKMNLTLEERHLDARVLAARVLYLKNPNVKITTDGLSLTLADKEKQAOAGKVPK 876	A.Molecule type: mRNA
QY 894 WMALESILRRFTHQSDVWYGVTVWELMTFGAKPYDGIIPAREIPDLLLEGGERLPQPIC 953	A.Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
DB 877 WMALESILQWYTHQSDVWYGVTVWELMTFGSKPYDGIIPAKETIASYLENGERLPQPIC 936	A.Cross-References: GB:M4309; NID:G183990; PIDN:AAA35979.1; PID:G306841
QY 954 TIDVYMWKCMMDISECRPRFRFELVSEFSMADPQRFVITQNEIDIGRASPILDSTFYNS 1013	C.Genetics:
DB 937 TIEVMIILKCMMDIPSSRPRFRFELVSEFSQADPBRVLIQO---NPLSLDRKLFGR 993	A.Gene: GDB:ERBB3; HER3
QY 1014 LLEDDMGDLVDAEYLVPOCGFCPPAPAGAGMVAHRRSSSTRGCGDLTLGEPSP 1073	A.Cross-References: GDB:119880; OMIM:190151
DB 994 LLSSDD--DVYDADBYLLPYKRI-----NRQGS----- 1019	A.Map position: 12q13-12q13
Query Match 34.8%; Score 2382.5; DB 2; Length 1342;	
Best Local Similarity 39.9%; Pred.No.1.38-90;	
Matches 525; Conservative 196; Mismatches 461; Indels 135; Gaps 32	
QY 10 GILLALLPPGAA--STQVCTGDMKLRLPASPEHLDMLRHLYOGQGVQGNLELTLP 67	11 GILTSLAGSEVNSQAVCPETLNGLSVTGDAEQVQTLKYLERCVVNGNLEIYLTGH 70
DB 11 GILTSLAGSEVNSQAVCPETLNGLSVTGDAEQVQTLKYLERCVVNGNLEIYLTGH 70	68 NASLSFLQDIEVQGVYLLAHNOVROVPLQRLRIVRGTLFEDNVALAVLDNCDPLNNTT 127
QY 68 NASLSFLQDIEVQGVYLLAHNOVROVPLQRLRIVRGTLFEDNVALAVLDNCDPLNNTT 127	DB 71 NADLSFLQWIREVYGVYLVANNESTLPLNLRVRYGTQYDDKFAIFVW-----LNTYNT 125
DB 71 NADLSFLQWIREVYGVYLVANNESTLPLNLRVRYGTQYDDKFAIFVW-----LNTYNT 125	QY 128 PYTGASPGGLRELOLRSTELILKGQVLIQNRPOLCYQDTILMKDIFHKNNQALLTLDTN 187
QY 128 PYTGASPGGLRELOLRSTELILKGQVLIQNRPOLCYQDTILMKDIFHKNNQALLTLDTN 187	DB 126 ----NSHARQLRLTLQTEILSGYVIEKDKXCHMDITDMWDIYDRD---AEIVVXD 178
DB 126 ----NSHARQLRLTLQTEILSGYVIEKDKXCHMDITDMWDIYDRD---AEIVVXD 178	QY 188 RSRACHPCSPMCKSRRCMGSESSDCSLRTFVQAGC-ARCKGPLPTDCHEQCAAGCTG 246
QY 188 RSRACHPCSPMCKSRRCMGSESSDCSLRTFVQAGC-ARCKGPLPTDCHEQCAAGCTG 246	DB 179 NGRSCPRHEVCCK-KCMGPGSEBCQTLTITICAPQCNMGCFPPNPNQCHDCACGCGS 237
DB 179 NGRSCPRHEVCCK-KCMGPGSEBCQTLTITICAPQCNMGCFPPNPNQCHDCACGCGS 237	QY 247 PKHSDCLACHFNHSGICEIHCPLVTTNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
QY 247 PKHSDCLACHFNHSGICEIHCPLVTTNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306	


```

Db      238 PDDTDFACGHEFNDGACVPRCPQLVYNNKLTFLQEPNPHTKYQYGVASCPHFNFV-V 296
Qy      307 DVGSCITLVGRLNHQETAEDGTQRCCKSKPCARVCYGLGMNNTVFWLRVPKYSKH 366
Db      297 DQTSVRAQCPDMEVD-KNGLKMCPCGGLCPKACEGSGSSRFQT-----VDSN 347
Qy      367 LEEFAGCKKIFGSLAFPESEFDQDPAASNTAPQYIYANSKFITGLTYLYISAWPDL 426
Db      348 IDGFVACTKLGNLDPLITGLNDPMHKKI PALDEPLNFRVREITGLVNIQSPHMH 407
Qy      427 DLSPQNLQVIRRIILHNGAYS-LTLQGLIGISWLGIRSLRELGSLALIHNTLCVHT 485
Db      408 NFSVFSNLTTIGRSLYNRSFLIMKNLVLSLGRSLKEISAGRIYISARQLCYHNS 467
Qy      486 VPMDQTFRNPHQALLHTA-NRPEDECVGEGLAHQICARGHCWGPPTQVNCOSQFLGQ 544
Db      468 LNWTKYLRGTEERLDIKNNRRPCDVAAEGKVCDDLPCSSGGCWMGPQGLSCRYSHGG 527
Qy      545 ECVBECKVLOGLPREYVNAHCLPCHECOPONGSYTCGPADQCVAAHYKQPPCVA 604
Db      528 VCETHCNFLNGEPRERFAHECFSCHECOPMEGATCGSGSDTCACCAHFRDPHCVS 587
Qy      605 RCPGKVPDLSTYMPIMKFPDEGACOPDINCTHSCVDLDDKGPABQRA-----SPLTSI 660
Db      588 SCHEGVG--AKGPYKYPDVQNECPCHENCTQCKGPELDDCGLVLIKTHLTA 645
Qy      661 VSAVVGILLVVLGVVFGILIKRQOKR-KYTMRLLOETELVEPLTPSGAMPQOQMR 719
Db      646 LTVIAG--LVVIFPMIGGTFELVWRGRRIQNKRAMRYLERGESIEPLDS-BKANKYLAR 702
Qy      720 ILKETELRKVKVLGSGAFGVYKGIWIPDEWVKI PVALKYLRNTSKAKKEIDEXYV 779
Db      703 IFKETELRKLVGSGVFGTVHGWIPGESIKIPVCIKVIEDSGSQSQAIVTDHMLA 762
Qy      780 MAGVSPVYSRLIGILTSTVQLTQMLPYGCLLDHVENRGLSQDLIMWCMQIAKM 839
Db      763 IGLDHAHIVRLIGLPGSSLIQVLTQYPLGSLDHRGALGRQLLWMMQVIAKM 822
Qy      840 SYLEDVRLVHDLAANNVLYKSPNNKIDFGALRLDIDETYNADGKVPYIKMALES 899
Db      823 YLLEHGHWHRNLAAANNVLLKSPQVQVADFGVADLLPPDDKQLYSBAKPIKMALES 882
Qy      900 ILRRFTQSDVSYGVTVWELMTFGAKPYDGI PAIREIDLEKEGRLOPPICITIDVM 959
Db      883 IHFGKXTHQSDVWSYGVTVWELMTFGAEYAGLRRLAEVLDLEKEERLAQOICTIDVM 942
Qy      960 IMYKCMIDSECPREPRELVEFSRYARDPQRFVVIQNEULGPA--SPLDSTFYSLE 1016
Db      943 VMYKCMIDENIRPTEKELANETRMARDPPRYLVIKRES-GPGIAPGPBPHGLTNKLE 1001
Qy      1017 DDDMGDLVAEEVLVPOGFFCPDPAAGAGVWHRHSSSTRSGGDLTLGLEP-SEE 1075
Db      1002 EVELPELDLDLLEAE-----NLATTLTGSLSLPVGLTNPR 1042
Qy      1076 APPSPAPSEAGSDVFDGLGMAAKGLQSLPTH-PSPLORYSEDPVLVLP-----S 1128
Db      1043 GSGSLSPSSGY-MPMNQNLGSCQESAVSSSSRCPVSLH-----PMRPGCLASE 1095
Qy      1129 ETDGVA-----PLTCSQPR-----YNNQPRVQPPSPRGP----- 1162
Db      1096 SSEGHVTGSEALQEVSMCRSRSRSPRPRGDSAYHSQRHSLITPTPLSPGLEED 1155
Qy      1163 -----LPAARPAATLERAKTLP-GKNGV-----KDVAFGAGAVENPEYLTPOGAA 1210
Db      1156 VNGYVMDTHLKTGPESREGTLLSVGLSLGTEBED-----EYVNNRRRHS 1207
Qy      1211 PCHPPAPSPAFDNLVYWD-----QDPREGAPRSTKGTPTMENBYL 1255
Db      1208 P-PHPPPSLEELGYEYMDVSGDSLASLGTSQSCPLHPVIMPTAGTTPDEDEYM 1263

```

```

epidermal growth factor receptor homolog precursor - rat
N/Alternate names: ErbB3 protein; HER3 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C/Accession: J04387
R/Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sterke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant
A/Reference number: J04387; MUID:96096535; PMID:8522190
A/Accession: J04387
A/Molecule type: mRNA
A/Residues: 1-1339 <HEL>
A/Cross-references: GB:U29339; NID:9915389; PID:9915390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT for
C/Comment: This protein is a functional heregulin receptor that transduces si-
C/Genetic:
A/Genes: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein ki-
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-139/Product: epidermal growth factor homolog #status predicted <MAT>
F/640-659/Domain: transmembrane #status predicted <TM>
F/705-970/Domain: protein kinase homology <KIN>
F/713-721/Region: protein kinase ATP-binding motif
F/939,1051,1156,1196,1219,1257,1259,1273,1286,1325/Binding site: phospho-
Query Match 33.5%; Score 2294.5; DB 2; Length 1339;
Best Local Similarity 40.0%; Pred. No. 5, Le-87;
Matches 515; Conservative 176; Mismatches 437; Indels 161; Gaps 3
Qy      3 LAALCRWGLLLALPPGA--STQVCTGTDMKRLPASPETHLDMKRLHYQSGQVYVGN 59
Db      7 LQVLC---FLTLARGSBMGNSQAVCCGTLNGLSVTDADNQVQTLVKLYKECEVWGN 62
Qy      60 LETTYPTNASEPDIQEOGVYLLAHNQVQVPLQRLATVGTQVFEENYALAVLDN 11
Db      63 LETVLTGHNAADSLFQWIREVYAVLVVAMNERSVPLFNLVVGTQVYDGKFAIPV-- 12
Qy      120 GPLNNTPTVGTASPGSLREIQLRSLEILKGVLLIQNPQCYODTILMDIFHKNNQL 17
Db      121 ---LVNT---NSHALRQLKFTGLTEILSGVYIEKNDLCHMDITDMDIYVR-- 17
Qy      180 ALTLIDTNRSAQHCPSCPKGSRGWSSSEDCQSLRTTVAGGC-ARCKPRLPTDCHE 23
Db      171 GAEIVKKNAGNCPCHVEYCKG-KCMGGGPDDCQILTTITCAPQCNCGCFGPNPNQCHD 22
Qy      239 QCAAGCTGKSHDCLACHFNHSGICEHLCPALVTVYNDTFESMPNPGRYTFGASCVTA 29
Db      230 ECAAGCGSPDIDCFACRRFNDSGACVRCBPVLYNKLTQLQLEPNPHTKYQYGVVYAS 28
Qy      299 CPYNYLSTDVGSCITLVCPILHNOEYTAEDGTQRCCKSKPCARVCYGLGMFNFTVSEWLR 351
Db      290 CFHNFV-VDQTFVRAQCPDMEVD-KNGLKMCPCGGLCPKACEGSGSS-----R 33;
Qy      359 VKVYASHLSEEPAGKKTFGSLAFPESEFDQDPAASNTAPQYIYANSKFITGLTYLY 411
Db      340 YQTVDSNIDGVNCTKLGNLDFLITGLANDPMHKKI PALDEPLNFRVREITGLVNI 39;
Qy      419 SAMPSLPLDSYFQNLQVIRRIILHNGAYS-LTLQGLIGISWLGIRSLRELGSLALIHN 47
Db      400 QSMPPHMHNFVSFNLTTIGRSLYNRSFLIMKNLVLSLGRSLKEISAGRIYISAN 45;
Qy      478 TILCFVHYPMQLFRNHQALLHTA-NRPEDECVGEGLAHQICARGHCWGPPTQVNC 536
Db      460 QQLCYHNSLMTWRLIRGSEERLDIKYDRPLGECGLAEKVCDDLPCSSGGCWMGPQGLS 516
Qy      537 CSQFLRGCEVBECKVLOGLPREYVNAHCLPCHECOPONGSYTCGPADQCVAAHYK 596
Db      520 GKNYSREGVCTHGNFLGGEPRERFAHECFSCHECOPMEGATCGSGSDACCAHFR 575
Qy      597 KDPFCVARGSPGKVPDLSTYMPIMKFPDEGACOPDINCTHSC--VDLDDKGPABQRA 654

```

Db 580 RDGPHCVNSCPHIGL- AKGIYKYPDAQNECRPCHECTGCGNPGELDQCGAEVLM 637

QY 655 SPLTISAVAGLILVVLGVGVGLIKRROOKIR-KYTMRLLOTELVEPLPTSGAMP 713

Db 638 SKPHVIAVTVG--LAVIILMILGSGFLYKRGRIQKAMKMYLERGESIEPLDPS-EKA 694

QY 714 NQAKMRLKETELRKVKVLGSGAFGTVYKGMIPDGENVKIPVAIKVIRENTSPANKKEI 773

Db 695 NKYLARIFKETELRKVKVLGSGVGFVHKGIWIPGESIKIPVCIKVIEDKSGRQSPQAV 754

QY 774 LDEAYMAGVGSFYVRLLIGICLTSTVOLVTLQMPGCLLDHVRNRRGLSGODLLNKM 833

Db 755 TDMLAVGSLDHAHVRLLGLCGSSQLVLYQLPGLSLDVKQKRETLGQLLHNGV 814

QY 834 QIAKMSYLEDVLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHYHAGGKVPK 893

Db 815 QIAKMYLEHSMVARDLALRVMLKSPSOYQVADFGVADLLPPDDKQLHSEAKTPIK 874

QY 894 WMALESLRRFTHQSDVWSYGVTVLMTFGAKRPDGIIPAREIDDLLEKGRLLPQPIIC 953

Db 875 WMALESLHFGKYTHQSDVWSYGVTVLMTFGAEPAAGRLAEIDDLLEKGRLLAQPQIC 934

QY 954 TIDVYIMVKCMWIDSECRPFRELVSFESRMADPPQRFVVIQNEIDLGPASFLDSTFYS 1013

Db 935 TIDVYIMVKCMWIDENINPTFEELANETPRADPPRLVTKRAS-GRGT--PAAES 991

QY 1014 LEEDDMKGLVDAEYLVPOQGFCDPPAPGAGMVRHRRSSSTRSGGDTLGLPSE 1073

Db 992 VLTTEL--QEALEPEL-----DLDLLEAE 1017

QY 1074 E-----EAPRSPAPSEG-----AGSDVPDGLDGMGAKLQ 1105

Db 1018 EGLATSLGSLSLPTGTLTRPGSQLSPSSGVPKMANOSLGEACLDLAVLGGREGSR 1077

QY 1106 SLETHDPSPLQRYSEDPVLPSEITDGYV---APL-----TC-----SPOE---Y 1145

Db 1078 PISLH-PIPRGR-----PASESSGHVTSGEAEIQEKVSVCFSRSRSPRPGRDSAY 1129

QY 1146 VNPDPVRPQPPSPREGP-----LPARPPGATLEPAKTLSP-QKRGV----- 1187

Db 1130 HSGRHSLLPTVPLSPGLLEBEDNGYVMPDTHLNGASSREGTLSSVGLSTGTEED 1189

QY 1188 KDVFAGCAVENPEYLLPQGAAPQPHPP 1216

Db 1190 ED-----EEYEVNKRKRQSP-PRPP 1209

RESULT 11

TVFVLV

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

C/Species: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999

C/Accession: B00643; A00643

R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M. Cell 41, 719-726, 1985

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and p12

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: B00643

A/Molecule type: mRNA

A/Residues: 1-698 <NHL>

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

P/1-6/Product: gag protein (fragment) #status predicted <GAG>

P/7-59/Product: env protein (fragment) #status predicted <ENV>

P/60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>

F/194-459/Domain: protein kinase homology <KIN>

F/202-210/Region: protein kinase ATP-binding motif

F/229/Active site: Lys #status predicted

Query Match 25 %; Score 1766.5; DB 1; Length 698;

Best Local Similarity 52.2%; Pred. No. 12e-65;

Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18

QY 584 GREADOCVCAHYKDPFVCARCPGSGKDLSSYPMIWEFDEGACQCPRLNCTHSCVDL 643

Db 60 GP--DHQMCAPHIDGPHCVKACPAVLGENDTL-VKXADANAACQLCHPCTGCKGP 116

QY 644 DNGCCAEQASPLTISAVV-GILVYVGLVGVGLIKRQOKIRKYTMRLLOETEL 702

Db 117 GLEGC--NGSKTPSIAGVVGGLCLVVLGILVYRR-HIVRKRTLRLLQERL 172

QY 703 VEDLTPSGAMPNQAQRILKETELRKVKVLGSGAFGTVYKGMIPDGENVKIPVAIKVLR 762

Db 173 VEDLTPSGAMPNQAQRILKETELRKVKVLGSGAFGTVYKGMIPDGENVKIPVAIKVLR 232

QY 763 ENTSPANKKEILDEAYVMAGVSPVYSLIGICLTSTVOLVTLQMPYGCCLDHVRNRRGR 822

Db 233 EATSPANKKEILDEAYVMAGVSPVYSLIGICLTSTVOLVTLQMPYGCCLDIYREHKN 292

QY 823 LGSQDLLNCOMQIAKMSYLEDVLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETE 882

Db 293 TGSQYLLNCOMQIAKMSYLEDVLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETE 352

QY 883 YHAGGKVPKIMMALESILRRRFTHQSDVWSYGVTVLMTFGAKRPDGIIPAREIDDLLE 942

Db 353 YHAGGKVPKIMMALESILRRRFTHQSDVWSYGVTVLMTFGAKRPDGIIPAREIDDLLE 412

QY 943 KGRLLPQPICTIDVYIMVKCMWIDSECRPFRELVSFESRMADPPQRFVVIQ-NEDLG 100

Db 413 KGRLLPQPICTIDVYIMVKCMWIDSECRPFRELVSFESRMADPPQRFVVIQ-NEDLG 472

QY 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCDPPAPGAGMVRHRRSSSTRSG 106

Db 473 LPSPTSKRYRLTMEEDMEDLVDAEYLVPOQGF-----NSPT--- 513

QY 1062 GGDULTGLPSEEAERSPL-----APSEGASDVFDGLDGMGAKLQSLPTHDPSPLQ 111

Db 514 -----SRPLSLSATSNNSATCID-----RNGQHPVREDSTVQ 550

QY 1117 RYSEDPVLPSEET--DGVVAPLTCSPQEVYNNQDVAPQPPSPREGFLPAARPGATLE 117

Db 551 RYSEDPVLPSEET--DGVVAPLTCSPQEVYNNQDVAPQPPSPREGFLPAARPGATLE 585

QY 1175 RAKTLSPGKGVYKDV-----AFGCAVENPEYLLPQGAAPQPHPPAF 121

Db 586 ----TAMVQNGIYNNISLTAISKLPMDSRYQNSHTAVDNBYL-----NNQSPFLA 633

QY 1220 SRAFDNLYWQ-----DPE-----RGAPPSFFKGTPTAENPEYGLDVP 1260

Db 634 KTVFSSPYWISGNHQINLDPDYOQDLFNETKRNGLLKVPAPENPEYLRVAP 689

RESULT 12

TVYVH

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (src

C/Species: avian erythroblastosis virus

C/Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C/Accession: A00644; A38022

R/Kamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K. Cell 35, 71-78, 1983

A/Title: The erbB gene of avian erythroblastosis virus is a member of the src g

A/Reference number: A00644; MUID:84026539; PMID:6313229

A/Accession: A00644

A/Molecule type: DNA

A/Residues: 1-604 <YAM>

A/Cross-references: GB:K01216; NID:9209676; PIDN:AAA42400.1; PID:9209678

R/Debure, B.; Henry, C.; Benabisa, M.; Bissette, G.; Claverie, J.M.; Saule, S.; Science 224, 1456-1459, 1984

A/Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new

A/Reference number: A38022; MUID:84223957; PMID:6328658

A/Accession: A38022
 A/Molecule type: DNA
 A/Residues: 128, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
 A/Cross-References: GB:K02006
 C/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F.130-395/Domain: protein kinase homology <KIN>
 F.138-146/Region: protein kinase ATP-binding motif
 F.165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;

Best Local Similarity 52.2%; Pred. No. 4.2e-63;
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

QY 593 CAHYKDPFCVACRPSGVKPDLSYPIWKPRDEGACOPPCINCTHSCVDLDDKGGPAGQ 652
DB 3 CAHFIDGPHCVKACRPAVLGENDTL-VKRYADANAVQQLCHPCTRCCKGPGLEGCP--- 58
QY 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYMRLLOETELVEPLTPSGA 711
DB 59 NGSKTPSIAGVVGGLCLVVGIGLGLYLR-R-HIVKRTLRLLQERELVEPLTPSGE 117
QY 712 MPNOAKMILKETELRKVKVLGSAFGTYVKGIMIPGENVKIPVAIKVLRNTSPKANK 771
DB 118 APNOAHLRLKETELFFKVKVVGSAFGTYVKGIMIPGENVKIPVAIKELRENTSPKANK 177
QY 772 EILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMYPGCLLDHVENRQRLGASODLLNW 831
DB 178 EILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMYPGCLLDHVENRQRLGASODLLNW 237
QY 832 CMQIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFFGLARLLDIDETEVNADGKVP 891
DB 238 CVOIAKGNLYLEERLRLVHRDLAANVLVKTPOHVKITDFFGLARLLDIDETEVNADGKVP 297
QY 892 IKMALESILRRRTTHOSDVWSYGVTVWELMTGAKYDGI PARETPDLEKGERLPORP 951
DB 298 IKMALESILRRRTTHOSDVWSYGVTVWELMTGAKYDGI PARETPDLEKGERLPORP 357
QY 952 ICTIDVYIMVYKCMWIDSECRPRELVSFESRMAKDPORFVYIQ-NEDLGPAFLDSTF 1010
DB 358 ICTIDVYIMVYKCMWIDSECRPRELVSFESRMAKDPORFVYIQ-NEDLGPAFLDSTF 417
QY 1011 YRSLLEDODMGDLVDAAEYLVPOQGFCCPPRPAAGAMVHRRSSSTSGGDLTLGLE 1070
DB 418 YRSLLEDODMGDLVDAAEYLVPOQGFCCPPRPAAGAMVHRRSSSTSGGDLTLGLE 449
QY 1071 PSEBEAPRSPL-----APSEGAGSDVFDGDLGMAKGLQSLPTHPSPLOQRYSEDPVP 1125
DB 450 -----SRTPLSLSLATSNNSATNCID-----RNGQGHVREDSPVQRYSSDPFGN 495
QY 1126 LPSET--DGIYAPLTGSPQPEYVQPDVPRQPPSPREGPLPAARPAAGATLERAKTUSPK 1183
DB 496 FLESIDDGFL-----PAPRYNQ--LMPKKESTAN----- 524
QY 1184 NGVYKDVFAF-----GAVENPEYLTGCGAARPPHPPAPSPAPD 1224
DB 525 --VQNGIYNFISLAIKSLPMSDRYQNSHSTAINDPEYL-----NNQSLAKTIVE 574
QY 1225 NLVYWDQPPPERGAPSTFKGTPTAENPEY 1264
DB 575 SSPYWIQSGNHQ-----INLDNPDY 594

```

RESULT 13

S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C/Species: avian erythroblastosis virus
 C/Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C/Accession: S35745
 R/Vennstrom, B.
 submitted to the EMBL Data Library, March 1993

A/Reference number: S35743
 A/Accession: S35745
 A/Molecule type: DNA
 A/Residues: 1-544 <VEN>
 A/Cross-References: EMBL:X12707
 C/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-
 F.135-400/Domain: protein kinase homology <KIN>
 F.143-151/Region: protein kinase ATP-binding motif
 F.170/Active site: Lys #status predicted

Query Match 24.0%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 7.4e-61;
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

```

QY 584 GPEADQVCACAHYKDPFCVACRPSGVKPDLSYPIWKPRDEGACOPPCINCTHSCVDL 643
DB 1 GP--DHCMCAHFIDGPHCVKACRPAVLGENDTL-VKRYADANAVQQLCHPCTRCCKGPG 57
QY 644 DDKCPAAGASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYMRLLOETEL 702
DB 58 GLESCP---NGSKTPSIAGVVGGLCLVVGIGLGLYLR-R-HIVKRTLRLLQEREL 113
QY 703 VEPLTPSGAMPNOAKMILKETELRKVKVLGSAFGTYVKGIMIPGENVKIPVAIKVLR 762
DB 114 VEPLTPSGAMPNOAKMILKETELFFKVKVVGSAFGTYVKGIMIPGENVKIPVAIKELR 173
QY 763 ENTSPKANKILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMYPGCLLDHVENRQR 822
DB 174 ENTSPKANKILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMYPGCLLDHVENRQR 233
QY 823 LGSQDLNMCQIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFFGLARLLDIDETE 882
DB 234 LGSQDLNMCQIAKGNLYLEERLRLVHRDLAANVLVKTPOHVKITDFFGLARLLDIDETE 293
QY 883 YHADGVPIPKMALESILRRRTTHOSDVWSYGVTVWELMTGAKYDGI PARETPDLE 942
DB 294 YHADGVPIPKMALESILRRRTTHOSDVWSYGVTVWELMTGAKYDGI PARETPDLE 353
QY 943 KGERLPORPCITDVIYIMVYKCMWIDSECRPRELVSFESRMAKDPORFVYIQ-NEDLG 100
DB 354 KGERLPORPCITDVIYIMVYKCMWIDSECRPRELVSFESRMAKDPORFVYIQ-NEDLG 413
QY 1002 PASPLDSTFYRSLLEDODMGDLVDAAEYLVPOQGFCCPPRPAAGAMVHRRSSSTSG 106
DB 414 LPSPDTSKPYRTLMEEDMEDIVDAEYLVPOQGF-----NSPST--- 454
QY 1062 GGDITLGLFSEBEAPRSPL-----APSEGAGSDVFDGDLGMAKGLQSLPTHPSPLO 111
DB 455 -----SRTPLSLSLATSNNSATNCIDRNGG-----H----- 481
QY 1117 RYSEDPVPPLPSETDGVVAPLTGSPQPEYVQPDVPRQPPSPREGPLPAARPAAGAT-LE 117
DB 482 -----PVREDFGL-----PAPRYNQ--LMPKKESTAN----- 523
QY 1176 AKTISPGKNGVWDVFAFGAVENPEYL 1203
DB 524 LPIDSRYN-----SHSTAINDPEYL 544

```

RESULT 14

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastos-
 C/Species: avian erythroblastosis virus
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C/Accession: S00727
 R/Scottling, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A/Title: Common site of mutation in the erbB gene of avian erythroblastosis vi
 A/Reference number: S00727; MUID:88217326; PMID:2897102
 A/Accession: S00727

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MEUAAACRGGLLALLPFGA.....TFKGTPTANPEPYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6655	97.1	1355	1	ERB2_HUMAN
2	5859	85.5	1257	1	ERB2_RAT
3	5846.5	85.3	1254	1	ERB2_MESAU
4	3099	45.2	1210	1	EGFR_HUMAN
5	3076	44.9	1210	1	EGFR_MOUSE
6	2934.5	42.8	1308	1	ERB4_HUMAN
7	2915	42.5	1308	1	ERB4_RAT
8	2655.5	38.7	1167	1	XMRK_XIPMA
9	2391.5	34.9	1142	1	ERB3_HUMAN
10	2317.5	33.8	1139	1	ERB3_RAT
11	1935	28.2	1426	1	EGFR_DROME
12	1749.5	25.5	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIER
15	1558	22.7	703	1	EGFR_CHICK
16	1285	18.8	1323	1	L7R3_CAEEL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	733	10.7	1363	1	ILPR_BRALA
19	696	10.2	1382	1	INER_HUMAN
20	696	10.2	1382	1	INER_HUMAN
21	691	10.1	1383	1	INER_RAT
22	690.5	10.1	1372	1	INER_MOUSE
23	683	10.0	1370	1	INER_MOUSE
24	680	9.9	1477	1	HTK7_HYDAT
25	676	9.9	1297	1	IRR_HUMAN
26	668.5	9.8	1300	1	IRR_CAVPO
27	636	9.3	1367	1	IGIR_HUMAN
28	626	9.1	1390	1	INER_AEDAE
29	625	9.1	1373	1	IGIR_MOUSE
30	621.5	9.0	1370	1	IGIR_RAT
31	615	8.0	2146	1	INER_DROME
32	596.5	8.7	987	1	EPB4_HUMAN
33	588	8.6	1114	1	REF_HUMAN

34	585	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
35	583.5	8.5	984	1	EPB1_CHICK	O07494 gallus gall
36	581.5	8.5	977	1	EPB2_MOUSE	O03145 mus musculu
37	580.5	8.5	987	1	EPB4_MOUSE	P54761 mus musculu
38	577	8.4	757	1	HT16_HYDAT	P53356 hydra atten
39	574.5	8.4	984	1	EPB1_RAT	P09759 rattus norv
40	573.5	8.4	1053	1	PAK1_CHICK	O00944 gallus gall
41	573	8.4	902	1	EPB8_XENLA	O91736 xenopus lae
42	569	8.3	1068	1	PAK1_XENLA	O91738 xenopus lae
43	568.5	8.3	984	1	EPB1_HUMAN	P54762 homo sapien
44	567.5	8.3	985	1	EPB4_XENLA	O91571 xenopus lae
45	563	8.2	1052	1	PAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ERB2_HUMAN STANDARD; PRT; 1255 AA.
ID ERB2_HUMAN AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath U., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Franke U., Levinson A., Ulrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 727-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
RN [5]
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMBIREGULIN.
RN CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
RN (POTENTIAL).
RN SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL curation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11767; AAA35808.1; -.
DR EMBL: M11761; AAA35808.1; JOINED.
DR EMBL: M11762; AAA35808.1; JOINED.
DR EMBL: M11763; AAA35808.1; JOINED.
DR EMBL: M11764; AAA35808.1; JOINED.
DR EMBL: M11765; AAA35808.1; JOINED.
DR EMBL: M11766; AAA35808.1; JOINED.
DR EMBL: M11730; AAA35808.1; -.
DR EMBL: M12036; AAA35828.1; -.
DR EMBL: X03363; CA427060.1; -.
DR PIR: A25491; A25491.
DR PIR: A24571; A24571.
DR HSSP: P11362; IFGK.
DR Genem: HGNC:3430; ERBB2.
DR MIM: 164870; -.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR InterPro: IPR004019; YLP motif.
DR Pfam: PF00069; Pkinase: 1.
DR Pfam: PF00757; Furin-like: 1.
DR Pfam: PF01030; Recep_L_domain: 2.
DR Pfam: PF02757; YLP: 2.
DR Pfam: PD000001; Euk_Pkinase: 1.
DR SMART: SM00261; FU: 3.
DR SMART: SM00219; Tyrc: 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP: 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR: 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM: 1.
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KM Tyrosinase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 227 BY SIMILARITY.
FT DISULFID 224 225 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 654 654 I -> V.
FT VARIANT 655 655 I -> V.
FT VARIANT 655 655 I -> V.
FT CONFLICT 1170 1170 /FTID=VAR_004078.
FT SEQUENCE 1255 AA; 137909 MM; 39E9D9DA04DC6962 CRC64;
SQ
Query Match 97.1%; Score 6655; DB 1; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
1 MELALCRMGILLALPPGAASSTQVCTGDMKRLPASPETHLDMRLHYGGGVVQGNL 60
1 MELALCRMGILLALPPGAASSTQVCTGDMKRLPASPETHLDMRLHYGGGVVQGNL 60
61 ELTYLPNASLSFLDIOEVQSVLLIAHQVROVPLQRLIYRGTLFEDNYALAVLNG 120
61 ELTYLPNASLSFLDIOEVQSVLLIAHQVROVPLQRLIYRGTLFEDNYALAVLNG 120
61 ELTYLPNASLSFLDIOEVQSVLLIAHQVROVPLQRLIYRGTLFEDNYALAVLNG 120
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEOC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEOC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEOC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEOC 240
241 AAGGTGRHSDCLACHFNHSGICEHCPALVYNTDTESMPNREGRYTFGASCTYACP 300
241 AAGGTGRHSDCLACHFNHSGICEHCPALVYNTDTESMPNREGRYTFGASCTYACP 300
241 AAGGTGRHSDCLACHFNHSGICEHCPALVYNTDTESMPNREGRYTFGASCTYACP 300
241 AAGGTGRHSDCLACHFNHSGICEHCPALVYNTDTESMPNREGRYTFGASCTYACP 300
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTQCEKSCPCARVCGGLMFPNFTVFLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTQCEKSCPCARVCGGLMFPNFTVFLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTQCEKSCPCARVCGGLMFPNFTVFLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTQCEKSCPCARVCGGLMFPNFTVFLR-V 359
360 PKYSASHLEFPACKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 419
360 PKYSASHLEFPACKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 419
360 PKYSASHLEFPACKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 419
360 PKYSASHLEFPACKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 419
354 RAYLSANIOEFACKCKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 413
354 RAYLSANIOEFACKCKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 413
354 RAYLSANIOEFACKCKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 413
354 RAYLSANIOEFACKCKIFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGYLYS 413
420 AMPPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 479
420 AMPPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 479
420 AMPPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 479
420 AMPPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 479
414 AMPDPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 473
414 AMPDPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 473
414 AMPDPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 473
414 AMPDPSLDELSPONLQVIRGRIILHNGAYSLTLOGISWLGRLSRLREGSGALILHNTH 473
480 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 539
480 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 539
480 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 539
480 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 539
474 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 533
474 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 533
474 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 533
474 LCPVHTVPMWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTQCVNSQ 533
540 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 599
540 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 599
540 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 599
540 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 599
534 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 593
534 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 593
534 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 593
534 FLNGQEVVEECRLQLPREYVNAARCLCHEQCQPNQNSVTCFPEADQCVACHYKDP 593
600 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
594 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVAPCPSPGKDELSPYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
660 IVANVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 719
660 IVANVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 719
660 IVANVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 719
660 IVANVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 719
654 IISAVVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 713
654 IISAVVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 713
654 IISAVVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 713
654 IISAVVIGILLVVLGVVFGILLIKRPOQKIRKTYMRLLQETLVEPLTPSGAMPNOQMR 713

```

QY 720 ILKETEARKVVLGSGAGFVYKGIWIPDGENVKIPVAIKVLENTSPXANKIIDEAVY 779
DB 714 ILKETELRKVVLGSGAGFVYKGIWIPDGENVKIPVAIKVLENTSPXANKIIDEAVY 773
QY 780 MAGVGSFVYVSHLIGICTSTYQVLTQMPYGCILDHYRENGRLSGQDILNMOMJAKM 839
DB 774 MAGVGSFVYVSHLIGICTSTYQVLTQMPYGCILDHYRENGRLSGQDILNMOMJAKM 833
QY 840 SYLEDVRLVHRDLAARVVLVKS PNHVKITDFGLARLLDIDETEFYHADGGKVPKMALES 899
DB 834 SYLEDVRLVHRDLAARVVLVKS PNHVKITDFGLARLLDIDETEFYHADGGKVPKMALES 893
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYM 959
DB 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYM 953
QY 960 IMVKCMWIDSECRFPRELVSSEFSRMARDQRFVYVIONEDLGASPLDSTFYRSLLEDD 1019
DB 954 IMVKCMWIDSECRFPRELVSSEFSRMARDQRFVYVIONEDLGASPLDSTFYRSLLEDD 1013
QY 1020 MGDVDAEYVLVPOQGFCEPDPAFAGAGVHHRHRSSTRSGGGDLTLGLEPSEEARPS 1079
DB 1014 MGDVDAEYVLVPOQGFCEPDPAFAGAGVHHRHRSSTRSGGGDLTLGLEPSEEARPS 1073
QY 1080 PLASEGAGSDVPFGDILGMAAKLOSLPTHDSPLORYSDEPTVPLBSETDGYAPLTC 1139
DB 1074 PLASEGAGSDVPFGDILGMAAKLOSLPTHDSPLORYSDEPTVPLBSETDGYAPLTC 1133
QY 1140 SPOPEYVNPQDVRPQPPSPREGPLPAARPAATLEBAKTISPKNVAVDVAFGAGAVEN 1199
DB 1134 SPOPEYVNPQDVRPQPPSPREGPLPAARPAATLEBAKTISPKNVAVDVAFGAGAVEN 1193
QY 1200 PEVYLTPOGGAAPQHPAPPAPSPAFDNLYWDQDPPERGAPOSTFKCTPTAENPEVYGLDV 1259
DB 1194 PEVYLTPOGGAAPQHPAPPAPSPAFDNLYWDQDPPERGAPOSTFKCTPTAENPEVYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 2
ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06434;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
CN ERB2 OR NEU
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RA MEDLINE=66118662; PubMed=3945311;
RA Bargman C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein."
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sclerotic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).

```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofte F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein."
RL EMO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY BGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALU_INIT.
DR PIR; A24562; TYRNTNU.
DR HSSP; P11362; IPRGX.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00066; kinase_1.
DR Pfam; PF00757; Furin-like_1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu_3.
DR SMART; SM00219; Tyrc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1..21
FT CHAIN 22..1257
FT DOMAIN 22..654
FT TRANSMEM 655..677
FT DOMAIN 678..1257
FT DOMAIN 159..369
FT DOMAIN 473..646
FT DOMAIN 722..989
FT NP_BIND 728..736
FT BINDING 755..755
FT ACT_SITE 847..847
FT DISULFID 196..205
FT DISULFID 200..213
FT DISULFID 221..228
FT DISULFID 225..236
FT DISULFID 237..245
FT DISULFID 241..253
FT DISULFID 256..265
FT DISULFID 269..296
FT DISULFID 300..312
FT DISULFID 316..332
FT DISULFID 335..339
FT DISULFID 513..522

```

```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 589 586 BY SIMILARITY.
FT DISULFID 598 602 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 644 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

```

Query Match 85.5%; Score 5859; DB 1; Length 1257;
 Best Local Similarity 85.8%; Pred. No. 6,5e-306;
 Matches 1084; Conservative 58; Mismatches 112; Indels 10; Gaps 4;

```

QY 1 MELALCRWGLLALLPPGASTQVCTCTDMKRLPASPEHLDMLHLYOGQVYQGNL 60
DB 1 MELAMCRWGLLALLPPGIAGTQVCTDMKRLPASPEHLDMLHLYOGQVYQGNL 60
QY 61 ELTYLPTNASLFLDIOEVGYVLIANQVROYPLRLRIVRGSTQLFEDNVALAVDNG 120
DB 61 ELTYVPAVASLFLDIOEVGYVLIANQVRYPLRLRIVRGSTQLFEDKVALAVDNR 120
QY 121 DELNNTTPTV-GASGGIRELQLSLTILKGVLIQNPOLCYQDTILKDIIPHKNOL 179
DB 121 DPQNVASTPGRTEGRLRELQLSLTILKGVLIQNPOLCYQDMVLMKDYFRKNOL 180
QY 180 ALLTIDNRSRACHPSCPMCKSRCKSGESSEDCOSLRRTVAGGACAKGRLPTDCHEQ 239
DB 181 AFVIDIDNRSRACHPSCAKCKDNHCKGSPEDCOILITICTSCACAKCKRLPTDCHEQ 240
QY 240 CAAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDTPESMNPGRYTFGASCTYAC 299
DB 241 CAAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDTPESMNPGRYTFGASCTYAC 300
QY 300 PNYLSTDVGSCTLVCPAHNOEVAEDGTQRCCKSCPCARVCYGLGMNFVSPRLRV 359
DB 301 PNYLSTEVGSCTLVCPAHNOEVAEDGTQRCCKSCPCARVCYGLGMNFVSPRLRV 353
QY 360 PK-VASHLLEFPAGCKKIFGSLAPLPESFDQDPASNTAPQYIKANSKFIGITELTGLYI 418
DB 364 ARAITSDNVQEFDDCKIFGSLAPLPESFDQDPSSGAPLRPEQLQVFEITLLEITGLYI 413
QY 419 SAMPSLPLDLSVFQNLQYIRGRILHNKAYSLTLOGLSIMGLRSLPELSSGALIHNT 478
DB 414 SAMPSLPLDLSVFQNLQYIRGRILHNKAYSLTLOGLSIMGLRSLPELSSGALIHNT 473
QY 479 HLCFVHTVPMQDLFNNPQALHNTANREDE-CVGEGLACIQCACARHCGMPPTOCVNC 537
DB 474 HLCFVHTVPMQDLFNNPQALHNTANREDE-CVGEGLACIQCACARHCGMPPTOCVNC 533
QY 538 SQFLRGQCEVCEKRLQGLPREYVNAHCLPCHCECPQNGSVTCGPEPADQCAAHYK 597
DB 534 SHFLRGQCEVCEKRLQGLPREYVNAHCLPCHCECPQNGSVTCGPEPADQCAAHYK 593
QY 598 DPFCVACPSGVAPDLSYMPIMKFPDEGACQPCPINTCHSCYVDLDDCKCAPQORASPL 657
DB 594 DSSCVACPSGVAPDLSYMPIMKFPDEGACQPCPINTCHSCYVDLDDCKCAPQORASPL 653
QY 658 TSISAVVGIILVVVLGVVFIILKRRQOKRKYTMRLIQETELVEPLTPSGAMPNOQ 717
DB 654 TFIATVGVVFIILVVVLGVVFIILKRRQOKRKYTMRLIQETELVEPLTPSGAMPNOQ 713
QY 718 MRLKETELRKVKVLGSGAFGTVYKGIWIPDGEVVKIPVAIXVLRNTSPKANKELIDEA 777

```

```

DB 714 MRLKETELRKVKVLGSGAFGTVYKGIWIPDGEVVKIPVAIXVLRNTSPKANKELIDEA 773
QY 778 YMAVGSPYVSRLLIGLCTSTVOLVQMLPYGCLDHYENRGRGLSQDLNMCQIAK 833
DB 774 YMAVGSPYVSRLLIGLCTSTVOLVQMLPYGCLDHYENRGRGLSQDLNMCQIAK 833
QY 838 GMSYEDVRLVHRLDAAARNVYKPNVVKITDFGLARLLIDETEVADGAKYPIKMA 897
DB 834 GMSYEDVRLVHRLDAAARNVYKPNVVKITDFGLARLLIDETEVADGAKYPIKMA 893
QY 898 ESILRRRFTHSGDVSYGVTWELMTGAPRYDGIIPAREIPDLLEKGERLPQPPCTIDV 957
DB 894 ESILRRRFTHSGDVSYGVTWELMTGAPRYDGIIPAREIPDLLEKGERLPQPPCTIDV 953
QY 958 YMIWCKMIDSECRPRPRELVSEFSMARDPQRFVYIQNEDIQPSPLUSTYRSLLED 101
DB 954 YMIWCKMIDSECRPRPRELVSEFSMARDPQRFVYIQNEDIQPSPLUSTYRSLLED 101
QY 1018 DDMGDLVDAEELVPOQFPCCDPAPAGGVVHRRSSSTRSGDGLTGLPSEBEAP 107
DB 1014 DDMGDLVDAEELVPOQFPCCDPAPAGGVVHRRSSSTRSGDGLTGLPSEBEAP 107
QY 1078 RSLPAPSGASDVFPDGLGMAKGIQSLPTHDPSPLOYSQDPTVPLPSETDGYAPL 113
DB 1074 RSLPAPSGASDVFPDGLGMAKGIQSLPTHDPSPLOYSQDPTVPLPSETDGYAPL 113
QY 1138 TCSPOPEVYNQDVPPOPPSPREGPLPAABPAGATLERATLSFGKNGVYKDYAFAGAV 115
DB 1134 ACSPOPEVYNQDVPPOPPSPREGPLPAABPAGATLERATLSFGKNGVYKDYAFAGAV 115
QY 1194 ENPEYLVFREGTASPPHSPAFSPFNLYWQNSSEQPPSPNFEQTPAENPEYGL 125
DB 1194 ENPEYLVFREGTASPPHSPAFSPFNLYWQNSSEQPPSPNFEQTPAENPEYGL 125
QY 1258 DVPV 1261
DB 1254 DVPV 1257

```

RESULT 3
 ER22_MESAU STANDARD; PRT; 1254 AA.
 ID ER22_MESAU
 AC 060533
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERBB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxId=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T, Ushijima T, Ishizaka Y, Nagao M, Arai M,
 RA Yamazaki Y, Ishikawa T;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL: D16295; BAA03801.1; -
 DR HSSP: P1362; IREK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SMO0261; FU; 3.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 653 675 POTENTIAL.
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 368 CYS-RICH.
 FT DOMAIN 472 644 CYS-RICH.
 FT NP_BIND 720 987 PROTEIN KINASE.
 FT BINDING 726 734 ATP (BY SIMILARITY).
 FT ACT_SITE 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.
 FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 584 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 623 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 658 658 V -> E (IN ONCOGENIC NEU).
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.3%; Score 5846.5; DB 1; Length 1254;
 Best Local Similarity 85.3%; Pred. No. 3e-105;
 Matches 1077; Conservative 69; Mismatches 107; Indels 9; Gaps 3

QY	1	MELALCRWGLLALPPGAASQVCTGTDMLRLPASPETHLMDRLHYOGCQVQSNL	60
DB	1	MELAAWGWGMLLALLSPGASGQVCTGTDMLRLPASPETHLMDRLHYOGCQVQSNL	60
QY	61	ELTYLPNNAISFLQDIQEOGVYVLLAHNOVQVLPQRIAYVGTGLFEDNYVLAVDNG	120
DB	61	ELTYLPNNAISFLQDIQEOGVYVLLAHNOVQVLPQRIAYVGTGLFEDNYVLAVDNR	120
QY	121	DLNNNTTPTVGASPGRLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFPKKNOLA	180
DB	121	DLNNNTTPTGRTREGRLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFPKKNOLA	180
QY	181	LTLIDTNRSPACHGCSMKGSRCKSGSSSDCQSLRTVAGGACACKGLPFDCCHEOC	240
DB	181	PVDIDTNRSPACHGCSMKGSRCKSGSSSDCQSLRTVAGGACACKGLPFDCCHEOC	240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFTESMNPGRYTFGASCTYACP	300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFTESMNPGRYTFGASCTYTCP	300
QY	301	YNYLSTVGSCTLVCPILHNOEVTAEADGTQCEKCSKPCARVCYGLMFPNFVSWLRVP	360
DB	301	YNYLSTVGSCTLVCPILHNOEVTAEADGTQCEKCSKPCARVCYGLMFPNFVSWLRVP	360
QY	361	K-VASHLKEPAGCKKIFGLSLAPFESFDDPDSNTPAQYIKANSFGIGLTGLYLYS	419
DB	361	K-VASHLKEPAGCKKIFGLSLAPFESFDDPDSNTPAQYIKANSFGIGLTGLYLYS	419
QY	420	AMPDPLDLSVFQNLQYIRGRILHNGAYSLTLQGLISWLGRLSLRELGSGLALIHNTHT	479
DB	420	AMPDPLDLSVFQNLQYIRGRILHNGAYSLTLQGLISWLGRLSLRELGSGLALIHNTHT	479
QY	474	LCPHATPMPQPLFNPQALHSHGNSPEECGLKDPACYPCLAHGHCWGPQVCNCHS	533
DB	474	LCPHATPMPQPLFNPQALHSHGNSPEECGLKDPACYPCLAHGHCWGPQVCNCHS	533
QY	540	FLRQGEVCEGRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAYKDP	599
DB	540	FLRQGEVCEGRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAYKDP	599
QY	594	PFVAPRCPSGVKPELSYMPIMKYPDEEGMCPICINCHSCVDLDRGCPAEGRASPAT	653
DB	594	PFVAPRCPSGVKPELSYMPIMKYPDEEGMCPICINCHSCVDLDRGCPAEGRASPAT	653
QY	720	ILKETELRKVKVLSGAGFTVYKGIWIPGSENVKIPVAIKVIRENTSPRANKIIDEAVY	779
DB	720	ILKETELRKVKVLSGAGFTVYKGIWIPGSENVKIPVAIKVIRENTSPRANKIIDEAVY	779
QY	774	MAGGSGYVYRLLGICLTSTVQVLTQMLPFGCLLDHVRHREGLSQDILNNCVOJAKQM	833
DB	774	MAGGSGYVYRLLGICLTSTVQVLTQMLPFGCLLDHVRHREGLSQDILNNCVOJAKQM	833
QY	840	SYLEDVAVLVRDLAARVAVLKSBNHKTIDFGARLLDIDETEHADGGKVPKIMNALES	899
DB	840	SYLEDVAVLVRDLAARVAVLKSBNHKTIDFGARLLDIDETEHADGGKVPKIMNALES	899
QY	900	ILRRRFHQSDVNSYGTWELMTFAKPYDGIIPAREIDOLEKGRLPQPICTIDVYM	959
DB	900	ILRRRFHQSDVNSYGTWELMTFAKPYDGIIPAREIDOLEKGRLPQPICTIDVYM	959
QY	960	IMYKMWIDEECPRELYSESRVARDPQRFVYVQNEIDLPAISLDSFTFRSLLEDD	101
DB	960	IMYKMWIDEECPRELYSESRVARDPQRFVYVQNEIDLPAISLDSFTFRSLLEDD	101
QY	954	IMYKMWIDEECPRELYSESRVARDPQRFVYVQNEIDLPAISLDSFTFRSLLEDD	101
DB	954	IMYKMWIDEECPRELYSESRVARDPQRFVYVQNEIDLPAISLDSFTFRSLLEDD	101

RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN (15)
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Hongger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN (16)
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN (17)
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731689;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-glycosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN (18)
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Inasaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN (19)
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X00588; CA025240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;

DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;

Query Match 45.2%; Score 3099; DB 1; Length 1210;
 Beet Local Similarity 49.3%; Pred. No. 2,1e-158;
 Matches 626; Conservative 174; Mismatch 359; Indels 112; Gaps 23

11 LLLALLPPGAA--STGYCTGDMKRLPASPETHLMDRLRYGGCCVQVQGNLETLPTN 68
 14 LLALCPASRLAEKXKCGSTSKLTQLGTFEDHFLSLQMGFNNECVLNLITTYQRN 73
 69 ASLSFLDIOEVQGVYLIANQVRQVPLRLRIVRTQGLFEDNYALAVLDNGDPLNNTTP 128
 74 YLSFLKTIQEVAGYVLIANTVERIFLEVLQIRGMVYENSYALAVLSNYD----- 126
 129 VTGSPGSLRELQRLSTELLKGVLIQRNQLCYODTILMKDIFKKNQALTLIDTR 188
 127 ---ANKTGLKELPRNQLQELHGAVRPSNNPALCNVESIQWRDIVSSDFLSNMSDFQNH 183
 189 SRACHPCSPMGKSGRCSSESDQSLTRTVCAAGCA-RCKGPLPTDCHEGCAACTGP 247
 184 LGSQCKDPCSPNGSCWGAEBNCQKLTIKICAQCSGRGRKSPSDCHNQCACCTGP 243
 248 KHSQCLALFPHNSGICELHCPALVLYNTDPEFSMNPBGRTYFGASCYACPYNYLSTD 307
 244 RESDCLCRKFRDENTKDCPCPLMLYNPPTYQMDVNPBGKYSFGATCVKCPRYNVVD 303
 308 VGSCTLVCPLHNOEVTAEFGDTCRCKSKPCARVCGLGMFNNFTVSPMLRVKVASNH 367
 304 HGSQVRAKCAADSYEM-EDQVRCKCKCBGRVAVNGIGI-GFKSL-----SIATN 356
 368 EEPAGCKKIFGSLAPFESFDGPASNTAPQYIKANSKFIGITELGYLYISAMPDLP 427
 357 KHFNCTSIGDHLILPVAFRGDSFHTPLDPEQLDLKTVKIRIGPLIQAMPENRD 416
 428 LSVQNLQVTRGRLINHGASLTLOGLGISMLGSLRLREGSLALIHNTHCFTHTVP 487
 417 LHAENLEIRGRTRKQGGQFSLAVSLNTSLGRLSKELISQDVLIISNKKLYCVANTIN 476
 488 WDLFRNPQALHTANRPDECVSGELACHQLCARHGMGPRPTQVCNCSQFLRQECV 547
 477 WKTLFGISGQKTIINRGNSCKAIQVCHALCSPEGCMGPRPCVCSRYNSRREV 536
 548 BEGCVLQGLPREVYNAHCLPCHPEQCPNGSVTCFPPADQCVAAHYKDPFVAACP 607
 537 DKCNLEGEPRFEVENSECICQHPCECLPQAMNITCTGRPDNCICQAHYIDGRHCVKCP 596
 608 SGVXPDLSTVPIKPPDEBGAQPCPINCCHSCVDLDDKGCFAEQASLTIVASVVG- 666
 597 AGVMGENTL-VKVIADAGHVCHLCPNCTYGGCTGGLGECPTNGKRP--SIATMGCA 653
 667 --LLVVVLGVVEGILIKRROKIRKTYRRLLOETLEVEPLTPSGAMPNQOMRLKET 724
 654 LLLLVVLAIGIG--LFMRBRHIVRKTIRLLQRELEVEPLTPSGEARNQMLILKET 710
 725 ELKRVYLGSGAGTYKGIINPDGSNVKIPAIKYLRENTSPKANKELLDAAVYAGVG 784
 711 EFKKIRVLSGAGTYKGLMIPGSKVXI PAVKELREATSPKANKELIDAAVYASVD 770
 785 SPVSRLLGLICTSTVQVLTQMLPYCLLDHRENRGRIGSODLLNMCQIAKGSYLED 844
 771 NPHVCRLGLICLSTVQVLTQMLPYCLLDHRENRGRIGSODLLNMCQIAKGSYLED 830
 845 VRLVHDLAARNVLVKSENVKITTDFGLARLLIDDETEYHADGKVPYIKMALSIILRR 904
 831 RRLVHADLAARNVLTPOHVITITFGKLAKLGAEEKEVHAEGKVPYIKMALSIILHRI 890
 905 FTQSDVMSYGVTVMEITFGAKPPDGI-PAREIPDLEKGERLPPPICTIDVYIMWKC 964
 891 YTHQSDVMSYGVTVMEITFGSKPIDGIPASISISILEKGERLPPPICTIDVYIMWKC 950


```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 128 128 (BY SIMILARITY).
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2F5 CRC64;

```

```

Query Match 44.9%; Score 3076; DB 1; Length 1210;
Best Local Similarity 49.2%; Pred. No. 3.6e-157;
Matches 630; Conservative 169; Mismatches 361; Indels 120; Gaps 26;

QY 11 LLLALLPPGAA--STQVCTGTDMKLRLPASPTHLDLRLHYGQCVQGNLELTLPFN 68
DB 14 LLTALCAAGALBEKKVCOGTSNRLLTGLTFEDHFLSLQRMVNCVGNLEITTYQGN 73
QY 69 ASLSFLDIOGVGVYLIANNOVQVLRRLVTRGTQFEDNYALAVLDNGDPLNTPP 128
DB 74 YDLSFLKTIQEVAVYLIANTVERIPLENIQIRGALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSTLTKKGVLIQRNPOLCYOTILMKCI-----PHKNQGLATLI 184
DB 125 -YGTNRIGLRELPMRNLOEILIGAVRSPNPILCMDDTIQMRDITVONVMSMSMDL--- 180
QY 185 DTNRSRACHPCSPKCGSRKSGSSSEDCQSLRTVAGGA--RCKGRLPTDCHEQCAAG 243
DB 181 -QSHPSSCPCKDPSCPGSCMGGEENCQKLTIKICAOQSHRCRGRSPSDCHQCAAG 239
QY 244 CTGPKHSDCLACHFNHNSGICELHCPALTYNTDTESMPNPEGRRTFGASCYTACPNY 303
DB 240 CTGPRESDCLVCKQFQDEATCKTCTPLMLYNTTYQMNVNPEGKYSFGATCVKCKPRY 299
QY 304 LSTDVSGSLVCPHNOEVAEDGTORCEKSKPCARVCGIQM--FNNFTVSFMLRVK 361
DB 300 VVTDHSGCVARACGPDYEV--EEDGIRKCKKCDGPKRKNVNGIGIGFRK-TIS----- 350
QY 362 VSASHLEEFAGCKKIFGSLAFLESDGDPASTAAQYIKANSKIFGITELTGYLYISAM 421
DB 351 INATNIGHFYKCYAISGDHILVAFKGSFTPTPLDRELEILKTYEITGFIILQAM 410
QY 422 PDSLPLDSVFQNTQVIRGLIHNGAVSLTQGLISMLGRLSILRELSGLALIHNTILC 481
DB 411 PDNMTDLAHFENLEIRGRKTKGCGSLAVGLNITSLGRLSKETISDGVYIISGRNVL 470
QY 482 FVATVPWQDLFRPHQALLHTANRPEDCEVGEGLACHQICARHGCMGPGPTQCVNCSQFL 541
DB 471 YANTINMKLFGFPNOKTKIMNRAEKDKAVNHNPLCSSEGCMPGPRDVCVCQVUS 530
QY 542 RGQECVECEGVLTQGLPREVYNAHGLRCHREQCPQNGSVTCGPRADQCVACAHKDPFF 601

```

```

DB 531 RGRCEKCNILLEGEPREPVNSBCIQCHECLPQANNTICTGRGPNDCIQCAHYIDGPH 590
QY 602 CVACPSGVKNDLSYMIWKFPPBEGACQCPINCHTSCYVDLDDKCPACRASPILTSIV 661
DB 591 CVKCPAGIMKENNTL-VMKYADANNVCHLCHANKCTYGCGPGLQCGEVPWSPGPKPISA 645
QY 662 SAVGILLVVLGVVFGI-LIKRQCKIRKTYRRLLOETELVEPTLTPSGAMNQAMRI 720
DB 650 TGIYGLLFTIV-VALDIGLPMRRRIHVRRTLRLLQSELVEPLTPSEARNOAHRI 708
QY 721 LKTELKRVKVLGSGAFGYKGIWPDGENVKI PVAIKYLRNTSPKANKELIDEAAYW 780
DB 709 LKTEFEKIKVLGSGAGTGYKGIWPEGEKVKI PVAIKELREATSPKANKELIDEAAYW 768
QY 781 AGVSPVSRLLGLCLSTVQVLTQMPVCCLDYHRENRGRIGSQDILNWCQIAKGS 840
DB 769 ASVNPVHCRLILCLSTVQVLTQMPVCCLDYHRENRGRIGSQDILNWCQIAKGN 828
QY 841 YLEDRVLVHRDLAARNVLPVSPNHVXITDFGLARLLDIDETEYHADGKVPIMMALESI 900
DB 829 YLEDRVLVHRDLAARNVLPVSPNHVXITDFGLARLLDIDETEYHADGKVPIMMALESI 888
QY 901 LREFTQSDVWSYGVYVWELMTFGAKPYGIPAREIPDLEKGERLPQPPITIDVYMI 960
DB 889 LHRITQSDVWSYGVYVWELMTFGSKPYDIPASDISISILEKGERLPQPPITIDVYMI 948
QY 961 MVKWMIMDSRPPREFRVLSEFSRMARDPQRFVIO-NEDLGASPLDSTFRSLLEDD 101
DB 949 MVKWMIMDASRPREFRLLIEFSKARDPQRYIVIGDERMHLPSPTDSNFRALNDEED 100
QY 1020 MGDVDAEVLVPOQGFCCPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEEARPS 107
DB 1009 MEDVDAEVLVPOQGF-----NSPST-----SRT 103
QY 1080 PLARSEGASDVDDGDLGMAAGLQSLPETHDSPIQRSEDPVLPSET--DGVAVL 113
DB 1035 PLSSLSATSN---NSTVACINRNGSCRYKEDAFQRYSSPTGAVTBDNDADFL--- 108
QY 1138 TCSPOPEYVQNPVRRPQPSPREGRPLPAARPAQATLERAKTISPGRNGVAKVFAFGAV 119
DB 1068 ---FVPEYVQ--SVPRKPAQSVGNPYVHNQPLHP-----APGDLHYQN--PHSNV 113
QY 1198 ENPEYLT-TPQCGAAPQHPPPAPSPAFNDLYWQ-----DF-----PERGAP 124
DB 1134 GNPEYLTQAQ-----PTLSSFSNPALMIQGHQMSLDNDYQDDFFPKETKN 118
QY 1241 STFGRTPAENPEYLGADV 1260
DB 1185 GIRKG-PTAENAEYLKVAAP 1203

```

RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 ID ERB4_HUMAN
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=838326;
 RA Plozman G.D., Cuioscu J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 RA Roy L., Neudauer M.G., Shoyab M.,
 RA "ligand-specific activation of HER4/p180erbB4, a fourth member of the

Query March	Best Local Similarity	44.7%	Score 2934.5	DB 1	Length 1308
Matched 605	Conservative 188	Mismatched 380	Indels 179	Gaps 258	
9	WGLTLLLPYGA	STQVCTGTDMKRLPASPETHLDMRHLVYQCGVYVQGNLEPTY	64		
8	WWWVSLVVAAGTVQPSDSQVCACTENKLSISLDSDEQYRALRKTYENCEVVMGNLEITS	67			
65	LPTNVAISLFIQDIQEVQGVYVLAHNVQVPLQRLIRVGTQLFEDNYALAVLNDGPELN	124			
68	IEHNRDLSFPRSVREVTGYVLAHQFRPLPLENRIIRGTKLVEDRYAVALAIPNVRKDG	127			
125	NTTPVPGASPGGLREIQLSLREILKGGVLIQORNPQLCYOQTLIMKQIFHKNQGLATLLI	184			
128	NP-----GLQELGKMLREILNGVYVYDQNFLLCYADLIHQDILVRNPWPSNLLTV	178			
185	DTNRSRACHPCSPMCKGSRQWSESSDQSLTRTVCAAGC-ARCKGPLPTDCCHEQCAAG	243			
179	STNGSSGCGCRCHKSCGTG-RCMQPTENHQTLIRTVCAQCGQRCGPGVYSDCHRECAAG	227			
244	CTGPKHSDDLACILHFPHSGICELHCPALVTYNTDFFSBMPREGRYTFGASCVTAACPYN	303			
238	CSGPKQTDPCACNFDNSGACVTCQCPQTFVNPFTFOLIEHFNNAKYTYGAFCKVCKCPHF	297			
304	LSLVDVGSCTLVCPRLHQEVTAAEDGQREKSKPCARVCYGLGFNNFTVSWFWLRVFKVS	363			
298	V-VDSSSCYRACRPSKMEV-EEENGIMKRCPTDLCPRKACDGI-----TGLSMASQTV	349			
364	ASHLEBPACCKIFGSLAFLPESFDGDPASNTAPQYIANKSKFGLTGLTYLISAMP	423			

```

Db      350 SSNIDKFINCTKINGNLFIIVTGIHGDYPYNAIEAIDBEKLNVEFTVEITGTINIQSWPP 409
Qy      424 SLPLDVSFQNLQVIRGRILHNGAYSLTLOGIGIMLGRSLREISGLALIHNTNHCYF 463
Db      410 NMTDFSVSNLVTIGRVLVYSGLSLILKOGGITSLOPQSKESASANIYITDNSNLCTY 469
Qy      484 HTVPWDQLFRNPQALHTANRPEDECVEGLACHOLCARCHWCMPGPTOCVANCQSGFLRG 543
Db      470 HTIMWTLFTSTINGIIVIRDRKKAENCTAEGMVCNHLCSDDGCGPDPDCLCRSRNRG 529
Qy      544 QECVECEVLOGLPREVYNARHCLPCHCECP-ONGSVTCGFPADQCVACATYKPPFC 602
Db      530 RLCIESCNLYGGEPEFENGSIYCECPQCKEMDGLTCHGPPDNCCKSHFKGPPNC 569
Qy      603 VARCPGVKPDLSYMPIMKFPDEEGACPCPCINCTHSCVDLDKGC-----PAE 651
Db      590 VEKCPDGLQGANF--IFKVADPDRCHPCHPNCTGCGNGTSHDCIYYWTHGSHLTPQH 647
Qy      652 QRASPLTIVSAVY-GILLVVVLGVVGIILIKRQOKIRKRYTKRRLLOETELVEPLTPSG 710
Db      648 AR-TPL--IAAGVIGGLFIVIGLTFVAVYRRKSIK-KKALRRFL-ETELVEPLTPSG 702
Qy      711 AMPNOAKRILKETELRKVKYLGSGAGFTYKGIWIDGENVKIPVAKYLENTSPKAN 770
Db      703 TAPNOAQRILKETELRKVKYLGSGAGFTYKGIWIDGENVKIPVAKILENTGPKAN 762
Qy      771 KEILDEAVMAGVGSYVYRLLIGLCTSTVOLVTLQMPYGLLDHVENRGRIGSQDPLN 830
Db      763 VEFMDALIMASMDPHLVRLIGVCLSTPTQLVTLQMPHGLLEYVHEKDNIGSQDPLN 822
Qy      831 WMOJAKGMSYLEDVRLVHRDLAARNVYKSPNNVKTITDRLAALDIDIDETVHAOGKY 850
Db      823 WCVQIAKGMVYLEERLVRDLAARNVYKSPNNVKTITDRLAALDIDIDETVHAOGKY 862
Qy      891 PIKMMALLESILRRPETHQSDVMSYGVTVWELMTGCAPIYDGIPIREIPDLLEKGERLP 950
Db      883 PIKMMALLESILRRPETHQSDVMSYGVTVWELMTGCAPIYDGIPIREIPDLLEKGERLP 942
Qy      951 PICTIDVTMIMYKCMIDSECRPRFRVSEFSMAADPORFVYIIONED-LGPASPLDST 1009
Db      943 PICTIDVTMIMYKCMIDSECRPRFRVSEFSMAADPORFVYIIONED-LGPASPLDST 1002
Qy      1010 FYRSLLEDGMDGLVDABEYLVPQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGL 1069
Db      1003 FQNLIDDEEDLEDMWDAEYLVP-QARNIIPP-----ITYSRKIDNSNS-----EIGH 1050
Qy      1070 EPSEEEAPRS-----PLAP-SEGAQSDVPDGLMGMA 1100
Db      1051 SPPPAYTPMSGNQFYVRDGGFAAEGVSVFYRAPFTSTIPKAPVAGATAEIPDSCCNGT 1110
Qy      1101 AKGLQSLPTHDPSPLOQRYSEDPVPLPS-----ETDGVYAPLTCSPQREYVNOQDVVR 1153
Db      1111 LRKPVAPVHVEDSSTQRYSDAPTFAPAPRSRGGLDDEGWTTPRDKPKQEVNHPVE-- 1167
Qy      1154 QPPSPREPLPAARPAATLERAKTLSPGKXGVKDVFAFGAVENPEYLTPOGGAAPQ 1213
Db      1168 -----ENPFSRR-----KNGDQ--ALDNEFYNAANG----- 1194
Qy      1214 HPPPA-----FSPAFDNTYYWDQDPPREGA- 1238
Db      1195 -PPKADEYVNEPLYLNTFANTLGAEYLNKNILSMPEKAKKADNDNDYVNHSLPPRSTL 1253
Qy      1239 -PPSTFKGTP-----AENPEYL 1255
Db      1254 QHPDYLOEYSTKIFYKQNGRIRPIVAENPEYL 1285

```

```

RESULT 7
ERR4_RAT
AC 062956: 0922N7; STANDARD; PRT, 1308 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
CN      ERB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON      NCBI_TaxID=10116;
RX      MEDLINE=9822155; PubMed=9553078;
RA      Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA      Matchioni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and Erb4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10266 (1998).
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Heart;
RX      MEDLINE=97184212; PubMed=91222560;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704 (1991).
RN      (2)
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704 (1991).
RN      (3)
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, Erb2 and
RT      Erb3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659 (1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NR-
CC      2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF041838; AAD08899.1; -.
CC      EMBL: U52531; AAC53051.1; -.
CC      HSSP: P11362; IFCK.
CC      InterPro: IPR000494; EGFR_L domain.
CC      InterPro: IPR000719; Euk_Pkinase.
CC      InterPro: IPR002174; Furin-like.
CC      InterPro: IPR001245; Tyr_pkinase.
CC      InterPro: IPR004019; YLP_motif.
CC      Pfam: PF00757; Furin-like; 1.
CC      Pfam: PF00069; Pkinase; 1.
CC      Pfam: PF01030; Recep_L_domain; 2.
CC      Pfam: PF02757; YLP_2.
CC      PRINTS: PRO0109; TYRKINASE.
CC      ProDom: PD000001; Euk_pkinase; 1.

```

DR SMART; SMO0261; FU; 4.
 DR SMART; SMO0219; TYK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT NP_BIND 724 732
 FT BINDING 751 751
 FT ACT_SITE 843 843
 FT DISULFID 189 197
 FT DISULFID 193 205
 FT DISULFID 213 221
 FT DISULFID 217 229
 FT DISULFID 230 238
 FT DISULFID 234 246
 FT DISULFID 249 258
 FT DISULFID 262 289
 FT DISULFID 293 304
 FT DISULFID 308 323
 FT DISULFID 326 330
 FT DISULFID 503 512
 FT DISULFID 507 520
 FT DISULFID 523 532
 FT DISULFID 536 552
 FT DISULFID 555 569
 FT DISULFID 569 577
 FT DISULFID 580 589
 FT DISULFID 593 614
 FT DISULFID 617 625
 FT DISULFID 621 633
 FT MOD_RES 1162 1162
 FT MOD_RES 1188 1188
 FT MOD_RES 1258 1258
 FT MOD_RES 1284 1284
 FT CARBOHYD 138 138
 FT CARBOHYD 174 174
 FT CARBOHYD 253 253
 FT CARBOHYD 358 358
 FT CARBOHYD 410 410
 FT CARBOHYD 473 473
 FT CARBOHYD 495 495
 FT CARBOHYD 548 548
 FT CARBOHYD 576 576
 FT CARBOHYD 620 620
 FT CONFLICT 1062 1062
 FT CONFLICT 1080 1082
 SQ SEQUENCE 1308 AA; 146957 MW; D944B8096A08B41 CRG64;

Query Match 42.5%; Score 2915; DB 1; Length 1308;
 Best Local Similarity 44.6%; Pred. No. 1.6e-148;
 Matches 603; Conservative 195; Mismatches 389; Indels 166; Gaps 29;

QY 1 MELA-ALCRMGILL--ALLPQASTGYCTGDMKRLRPAPFETHLMRLHVGCCVWG 57
 DB 1 MCLATGLWWSGLVAAKTVPQASQVSCAGTENTKLSLSLEQVYALKRYENGEVVM 60
 QY 58 GNLELTLYPTNASLSFLQDIOEVGYVLIANQVRQVPLQRLIVRGTOLEPDNYALAVL 117
 DB 61 GNLEITSIEHNEDLSFLRSIRKVTGYVALNORFYLPLENLRIRIRGKLYEDRYALAI 120
 QY 118 DNGDPLNNTTPTVTASPGGLRELRLSRSTELIKGVLILQNRPOLCQDTILMKDIFKKN 177
 DB 121 LNRKDGNF-----GLQELGLKNTLEILNGSVYVDQNKFLCYADTIHMQDIYANPW 171

QY 178 QLATLIDTRSRACHPCSPMKGSRGMBESSHSDCSLTRVYAGGC-ARCKGPLPTDCC 236
 DB 172 PSNMTLVSTIGSSGCGCRKSCCTG-RCWGTENHCQTLRLTVCAECCDRCCTGPPYSDCC 230
 QY 237 HEQCAAGCTPKKSHDCLAFHNSHGICELHCPALVTYNTDFESMPNDEGRYFGASCV 296
 DB 231 HREAGAGCGSPKIDTCACAKNPFDSGACTQCCTGVVYPTTFQLBHNFNKATYGAFCV 290
 QY 297 TACYNTLSTDVSGCTVCPCLHNOETAEADGTGRCKSKPCARVCYIGLGMFNFTVSM 356
 DB 291 KCPCHNFV-VDSSCVACAPSSKREV-ENGIMCKPCCTDICKPADGIG-----TGLS 342
 QY 357 LRVPEASHLEBPAGCKIFGSLAFIPESFDDPASPNTAPQYIKANSKFIGITELTVL 416
 DB 343 MSACTVSSNIDKINCTXKNGULIFLVGHIDDPNAIDAPLEKLNFRVREITGFL 402
 QY 417 YISAMPDLPDLSVFQNLQVIRGRILHNGAYSITLQGLISWLGSLBELSGLLIHH 476
 DB 403 NIQTMPPNMTDFSVFSMLVTIGRVLVYSGLSLTLIKQOGLTSLQFSLKEISAGNTYITD 462
 QY 477 NTHLCFHTVPMQLEFNPHQALHTANRPEDSCVEGLACHQLCARHGMCGPPTQCVN 536
 DB 463 NSNLCYHTHTMTTLFSTVQRIVIRDNRAENCTAEGMVCNHLCSNDCMGPGPQCCS 522
 QY 537 CSQFLRGQCEVEECRVLQGLPREVYNAHCLPCHPCQP-QNGSVTCFGEADQVACAH 595
 DB 523 CRFSPRGKICIESCNLYDGEFRENKSI-VEECDSCQCEKMEGDLTCHGPGPNCTKSH 582
 QY 596 YKDPFPCVAPSPSVKXKDLGYMPIMKPPDEGACQCPINCTHSQYDLDQK- 648
 DB 583 FKDPBNCEKPEVDLQANSF--IFKXADDRCHCPHPCVTCGCGPPTSHDCTIYPMWG 640
 QY 649 ----PAEGRASPLTISVAVV-GILLVVVGVFGLIKRQCKIRKYTRRLLOETELV 703
 DB 641 HSTLPQHAR-TPL--IAAGYIGLFIIVIALTFAYVVRKSK-K-KBALRRL-ETELV 695
 QY 704 EPLTPSGAMPQAOQRLIKETELKRVYVSSGAFYTYKIMIPDGENYVIPAIVYLE 763
 DB 696 EPLTPSGAPQAOQRLIKETELKRVYVSSGAFYTYKIMIPDGENYVIPAIVYLE 755
 QY 764 NTSPANKELIDEAYVWAGSPVYSLLGICTSTVOLTQAMPYGCGLDHYRNRGL 823
 DB 756 TTGPANVEFDEALIASVDHPLVLDVGLCSPTGLVTOIMPHCCLLEYHEKNDI 815
 QY 824 GSQDLNMCWIAKGSYLEDVRLVHDLAARVLYKSPNHVKITDFGLARLIDIDETEY 883
 DB 816 GSQDLNMCWIAKGMVYLERRLVHDLAARVLYKSPNHVKITDFGLARLEGEKEY 875
 QY 884 HADGKVPKIMALESLRRFTHQSVMSYGVTVMLTFGAKPYDGIPIAREIPDLEK 943
 DB 876 NADGKMPKIMALECHYKRFTHQSVMSYGVTVIMLTFGKPYDGIPIAREIPDLEK 935
 QY 944 GERLPQPICTIDVYIMVWKCMIDSECRPRFRELVSFARMARDPQRFVYIONED-LGP 100
 DB 936 GERLPQPICTIDVYIMVWKCMIDADSRPFRELAIEFSRMARDPQRFVYIOGDDRMKL 995
 QY 1003 ASPLDSTYRRLLEDDDMGDLVDAEELVYVQGGFFCDP----- 104
 DB 996 PSPNKSKEFQNLDEEDLEDDMDDEELVLP-QANIPPPYTSRTRIDRSRSITGHSPP 105
 QY 1042 --APGAGMVAHRRSRSSSTSGGDLTLGLEPSEEAAPRPLAPSEAGSDVDFDGLGMG 109
 DB 1055 AYTPHSSGFYVQGGFATQGG---MMPYATATSTIPAPFAV--QGATNEMPDSCNG 110
 QY 1100 AAKGIQSLPHDPSPLQRYSEDPIVPLPS-----ETDGYVAPLTSPPQPEVYNQDVR 115
 DB 1110 TLRKVPVHVVEDSSTORYSADPIVFAPEHRPALEDEEYMPHMDKPKQAEVLNVE-- 116
 QY 1153 PQPSPREGPLPAAPAGATLERAKTLSPGKNGVAVKVFAGGAVENFEVLTQGGAAAP 121
 DB 1168 -----ENPVSRR-----KNGDLQ-----ALDNPEVHSSSG----- 119
 QY 1213 PHPPPA-----FSPAFLNLYWDDQPPBERGA 123

Dd		1195	--PPEKAEVYNEPLYNFTNALGNAAEYMKNSLSYPEAKAKAFPNPDYMHSLPPRST	1253
Oy	1239	--PPTFKGTPT-----ANDEPYL	1255	
Dd		1253	LOHPDYLOEYSTKYFYKONGRIAPVAENPEYL	1285
 RESULT 8 XMRK_XIPMA STANDARD; PRT; 1167 AA.				
AC	P13388;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).			
GN	XMRK OR TU.			
OS	Xiphophorus maculatus (Southern platyfish).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;			
OC	Cyprinodontiformes; Poeciliidae; Xiphophorus.			
CC	NCBI_TaxID=8083;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90015140; PubMed=2797166;			
RA	Witbrodt J., Adam D., Mallschek B., Maueier W., Raulf F.,			
RA	Telling A., Robertson S.M., Schachtl M.;			
RT	"Novel putative receptor tyrosine kinase encoded by the melanoma-			
RL	Nature 341:415-421(1999)."			
RN	[2]			
RP	REVISION TO 515.			
RA	Schachtl M.;			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.			
CC	-1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.			
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.			
CC	-1- SIMILARITY: BELONGS TO THE EGFR RECEPTOR FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X16891; CAA34770.2; -			
DR	PIR; S06142; S06142.			
DR	HSP; P13362; IFGK.			
DR	InterPro; IPRO00494; EGFR_L domain.			
DR	InterPro; IPRO00719; Euk_kinase.			
DR	InterPro; IPRO02174; Furin-like.			
DR	InterPro; IPRO02290; Ser_th_kinase.			
DR	InterPro; IPRO01245; Tyr_kinase.			
DR	Pfam; PF00069; pkinase; 1.			
DR	Pfam; PF00757; Furin-like; 1.			
DR	Pfam; PF01030; Recep_L domain; 2.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	ProDom; PD000001; Tyr_kinase; 1.			
DR	SMART; SM00261; FU_5			
DR	SMART; SM00220; STKC; 1.			
DR	SMART; SM00219; TYKC; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.			
KM	Transmembrane; Glycoprotein; Receptor; Signal; Transferase;			
KM	Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.			
LT	SYNOPSIS			
LT	CHAIN 26 1167			
MELANOMA RECEPTOR PROTEIN-TYROSINE				

Query Match	Best Local Similarity	Score	DB	Length
Matches 569; Conservative 173; Mismatches 384; Indels 151; Gaps 315	44.6%	2655.5	16	167
4	AAICRMGLIALLP	PGCAST---	QVCTGTDMKRLPASPETHLDMRLHYQGVQVQGN	59
8	AAIIQ--LLVLISRCSCSTDPDRKVCQGTSTNQMTM--	LNHHYLMKKQWYSGCNVLE	62	
60	LELTYPPTNASTLSFLDDIOBVGCVYLIANNQVRQVPLORLRIYRGTOFLFEDNYALAVLDN	119		
63	LETTYQENODLSFLDSIQEAVGGVYLIANNEVSTPLVNLRLIRQNLYEENFTLLVMSN	122		
120	GDPLNTTPTVTCGASPGGLRELOJRSSTELTKGCVLIQNNPOLCYODTILMKDIFPKXNQL	178		
123	YQK-NPESP--DYQYGLKQLOJNSNITELSGVAVSHNPILLCVETINMMDIDYKTSNP	179		
180	ALTLITDNRRAACHPCSPMKCKSGRCCKGESSDDQSLRTVCAAGC-ARCKGLPTDCCHE	238		
180	TMLILPAPFERQOCQCDHGCNVGASGAPRGHCQKFTYLLTAEQCNRCRCRGPKEIDCNE	239		
239	QCAAGCTGPHGSCCLACLFHNSGICELHCPALVTYNDFESMNPREGRTTFCASCUTA	298		
240	HCAAGCTGPRATCTCLACRPFNDGTCKTCPCPKYIDIVSHQVYDNPRIKTYTFBACVKE	299		
299	CPVNYLSTDVGSCITLVCPLHNOEVTAAEDQTCRCKSPCARVCYGLGMFN-NFTVSFWL	357		
300	CPSNVYVTE-GACVCRSGSAGMLEV-ENGKSKCRPCDVCVKPCDGIIGSLNSTIA---	354		
358	RVRKVASHLIEBPAQCKRTFGSLALPESPQDP---ASNAPQYIANKSFGICITELTG	414		
355	---VNSTNRISNCTKINQDIILNRRNSFEDCPHYKIGTMDPEHL---NMLTVEITG	407		
415	YLKISAMPDLPDLVDFONTQVIRRIILHNGVSLTLQGLIGIMLGRSLRELGSGLAL	473		

KM Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1342
 FT DOMAIN 20 643
 FT TRANSMEM 644 664
 FT DOMAIN 665 1342
 FT DOMAIN 709 966
 FT NP_BIND 715 723
 FT BINDING 742 742
 FT ACT_SITE 834 834
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 552 565
 FT DISULFID 566 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT VARSPLIC 141 183
 FT VARSPLIC 184 1342
 FT CONFLICT 560 560
 FT CONFLICT 1064 1064
 SQ SEQUENCE 1342 AA; 148097 MM; 7201E7F6C3A374BD CRC64;
 Query Match 34.9%; Score 2391.5; DB 1; Length 1342;
 Best Local Similarity 39.9%; Pred. No. 1.6e-120;
 Matches 526; Conservative 196; Mismatches 460; Indels 135; Gaps 32;
 QY 10 GLLALLPFGAA--STQVCTGDMKRLPASPEFHLMLHLIYGGCCVVOGNELEYLPT 67
 DB 11 GLPFLARGEVSGNSQAVCPGTNLGLSTGDAENQYOTLVKYERCEVWNGNLEIYLTC 70
 QY 68 NASLFLQDICEVOGYVLIANQVROVPLQRLIRVSTQLEFENYALAVDNDPLNNT 127
 DB 71 NADSLFLQMIREVGYVLIANNESTLEPLRLRYRSTQYVDGKFAIYV-----LNYNT 125
 QY 128 PVTGASPGGLRELOLRSTETILKGGVLIQRNPQLCYODTILMKDIFKNNQLALTLDIN 187
 DB 126 -----NSSHALRQLTLQITELISGVYIEKDKLCHMDITIDMRDIYDRD---AETVMD 178
 QY 188 RSRACHPCSPWCKSRGSESEDCOSLTPTVCAAGC-AQCKGRLPTDCCHEGCACTG 246
 DB 179 NGRSCPPCHEYCKG-RCKMGPSSEDCQTLTKITCAPQCNGHCFGNPNQCCHDCAGCGSG 237
 QY 247 PKHSDCLACLHFNHSGICEHCPALVTYNTDTESMNPBGRATYFGASCVTACPYNYLST 306
 DB 238 PQTIDCFACRHFNDSGACVPRCPQLVYNKLTFLQLEBNPHTKYQYGVGVCAASCPHFV-V 296

QY 307 DVSGCTLVCPILHNOEVTAEQGTORCEKSKRCARVCYGLGFMNFNTVSFWLRYBKVASH 361
 DB 297 DQTSVCRACPPDKREVD-KNGLKXCEPCGGLCPACGCTGGSGFQ-----VDSN 34
 QY 367 LEFPAGCKKIGSLAFLEPSFDGSPASNTAPQYIKANSKFGITELTGILYISAMPDLP 421
 DB 348 IDGFNCKRIKIGNDLFLTGLNGPFWKXIPALDPEKLNVRFTVEITGYLNIQSWPHMH 40
 QY 427 DLSPQNLQVIRGRILHNGAYS-LTLOGLISWGLSLRLSGLLIHNTHLCFVHT 48
 DB 408 NPSVNSNLTTIGGSLNYRGSGSLIMKNLNTVLSGFLSKLSISGRYIISNRQLCHHS 46
 QY 486 VPMQDLFNPQIALIHTA-NRDECEYGEGLACHQLCARHGWPGSPPTQCVNCSQFLRG 54
 DB 468 LMTVTLRGPTEERLIDIKHNRPRRDCAEGVCDPLCSSGCGWPGGQCCLSCNYSRG 52
 QY 545 ECVEECRLQGLPREVYNARHCLCHEPCPONSQVCPFGEDAQCAAHYNDPRCVA 60
 DB 528 VCTHGNFLNGBEPPEFAHEACFSCHEPCQPMESTATCNSSGSTGCAQCAHFRDGPCHVS 58
 QY 605 RCPGSKEDLSYMPIWKFPEDEGACQCPINCTHSQVLDLDDKGPABQRA-----SPLTSI 66
 DB 588 SCPHGVLD-AKGPYKXPQDVQNECRCHENCTGCKGPELDQGLQTLVLIGKTHLTA 64
 QY 661 VSAVVGILLVVLGVVGLIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNQAOMR 71
 DB 646 LTVING--LVVI FMMLGSTFLYWRGRRIQWKRARRYLERGESIEPLDPS-EKANKYLAR 70
 QY 720 ILKETELKAVLVLSGAGFYVYKGIWIPQENYKI PAIKVLRNTPSPKANKELIDEAYV 77
 DB 703 IFKETELKLVLSGVGFYHKVWIPBESITKIPICIVIEKSRQSGFOAVTBMILA 76
 QY 780 MAGVSPYVSLIGICTSTVQVLTQMPYGCLLDHVRENKRGISGDLNMCQIAKGM 83
 DB 763 IGSLEHAIIVALLICGSSLSQVLTQYLPGLSLIDHVRQAGALPGLLNMVGQIAKGM 82
 QY 840 SYLEDEVLVHMDLAAVNVLYKSPHNVITPGLARLLDDETEHNDGCVPIKMALES 89
 DB 823 YLLEHGVAHNLAAVNVLYKSPQVAVDPAVADLLPPDCKLYLSEATPIKMALES 88
 QY 900 ILRRFTHQSDVMSYGVTVLWELMTFGAKPYDGIIPAREIPDLKEGERTPOPICTIDVYM 95
 DB 883 IHFGYTHQSDVMSYGVTVLWELMTFGAEVPAAGRLAEVPLLEKGERLAPQCTIDVYM 94
 QY 960 IMVCKMTIDSCRRFELVSESRMARDPQRFVIONEDLGRA---SPLDSTFYRSLE 101
 DB 943 VMVCKMTIDENIRPTFKELANEFTRMARDPRYLIVIRBS-GRGIAPGPEPHGLTNKLE 100
 QY 1017 DDMQGDVLDAEYLVVPOGFCPPAPAGAGAMVHHRSSSTSGGDLTLGLEP-SEEE 107
 DB 1002 EVELPELIDLDLLEAED-----NLATTLTGSALSLPVGTLMRPR 104
 QY 1076 APRSPPLABSEGAGSDVDPGLGMAANGLSLPTH-DPSPLQRYSEDPFVPLP-----S 112
 DB 1043 GSGQLLSPSSGY-PMNQGNLGECCQSAVSGSSERCPRVSIH-----PMRGCLASE 109
 QY 1129 ETDGVYA-----PLTSSPOPE-----YVNPQVDRPQPPSREBP----- 116
 DB 1096 SSBGHVGSSEALQEKYSKCRSSRSRSPRPDDSAVSHQSHLLTPVPLSPGLEED 115
 QY 1163 -----LPAAPAGATLERAKTISP-GNGGVV-----KQFAFGAVENDEYVLPQGA 121
 DB 1156 VNGVYVMDTHLKGTPSSREGLTSSVGLSVLGTBEED-----ELEYVNRARRRS 120
 QY 1211 PQHPPPAPFAFANLYYWD-----ODPBERGAPSTFFGTATENPEYL 125
 DB 1208 P-PHPPPSSLBELGVYMDVGSDLASLSIGSTQSCPLHPVPIPTAGTIPDDEYEM 1263
 RESULT 10
 ID_ERB3 RAT
 AC_Q62799; Q62955; STANDARD; PRT; 1339 AA.

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein".
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85; 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroregulin and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration".
 RL J. Neurosci. 17:1642-1659(1997).
 CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 CC -----
 DR EMBL: U29339; AAC08498.2; -
 DR EMBL: U52530; AAC3050.1; -
 DR HSSP: F11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk-kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr-kinase.
 DR Pfam: PRO0069; kinase.1.
 DR Pfam: PRO0757; Furin-like.1.
 DR Pfam: PRO1030; Recep_L_domain.2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk-kinase.1.
 DR SMART: SMO0261; FU; 5.
 DR SMART: SMO0219; TyrcK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 644 662 POTENTIAL.
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
 FT 183 259 CYS-RICH.
 FT DOMAIN 707 964 PROTEIN KINASE.
 FT NP BIND 713 721 ATP (BY SIMILARITY).
 FT BINDING 740 740 ATP (BY SIMILARITY).
 FT ACT SITE 832 832 BY SIMILARITY.
 FT DISULFID 186 194 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.
 FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 266 BY SIMILARITY.
 FT DISULFID 280 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CONFLICT 1088 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDPFI6 CRC64;
 Query Match 33.8%; Score 2317.5; DB 1; Length 1339;
 Best Local Similarity 40.2%; Pred. No. 1,4e-116;
 Matches 518; Conservative 176; Mismatches 434; Indels 161; Gaps 34
 QY 3 LAALCRWGLLALLPPGAA---STQVCTGDMKRLRPASETHLDMRLHYGCGOVQGN 59
 DB 7 LQVLC-----FLSLARGSEMNSQAVCPGTLNGSLVGDADNGVQTLKYKCEVVMGN 62
 QY 60 LEITLPTNASTFLQDIQEVGVLLAHNQVRVPLQRLIRYRGTOLEFDNYALAVLDN 119
 DB 63 LEIVTGHNADLSFLQWIREVTGYLVAMNEFSVPLPNLRVVRGVGVYDGKRAIFVM-- 120
 QY 120 GDPLNNTTPVTGASPGGIRELQSLTEILKGVLLQRPQLCYQDTILLKDI FHKNNQL 179
 DB 121 ---LNYNT---NSSHALRQLKFTQLTEILISGVYIEKNKLCHMDTILDRIVRR--- 170
 QY 180 ALTLIDTRSRACIPCPSPMCKGSCWSESSDQSLTRTYCAGGC-ARCGPIPTDCHE 238
 DB 171 GAETIVKKNGANCPCHVEVCKG-RCWEPGRDDCOILTKTICAPQCNCRGCPNPNOCHD 229
 QY 239 QCAAGCTGPKHSDGLACILFHNSGICGLHPALVTYVTDTFESMNPNEGRTYGASCVTA 288
 DB 230 ECAGGCGSPQDTDFACRRFNDSCACPRCPREPLVYTKLTFGLEPNPHTRYQVGVCVAS 289
 QY 299 CPNYVLSTDVSGCTLVCPLEHNOEYTAEDGTORCEKSKPCARVCYGIQMNFVTSFWMLR 358
 DB 290 CPHNFV-VDQTFVCVRACPPPKMEVD-THGDKMCEPCCGLCPKXACEGSGS-----R 339
 QY 359 VPKVSASHLEFACCKIFGSLAFPLPSFPGDPAASNAAPQYIKANSKFTIGITELTGTVLI 418
 DB 340 YQYDSSNIYDFVNVCTILGLNDELITGLNVDPHMKIPALDEPKLVNFTVREITGYLNI 399
 QY 419 SAMPDGSLPDLISVFQNLQVIRGRILHNGAYS-LTLQGLISWLGRLSLRELQGLALIHNN 477

```

Db      400 QSWPPHNHFVSFNSLTTIGRLSLYNGFSLIMKLANLTSIGFSLKEISAGRYISAN 459
Qy      478 THLCFVTFVPMDOLEFRPHQALLHTA-NRPEDCEVBEGLAQHQLCARHGWKPGTQVCN 536
Db      460 QQLCYHSHLMTLRLKPSSEERLDITKRLDGLCLBEGKYCDPLCGSGCWPGGQGLS 519
Qy      537 CSQFLRGQECVEECRYLQGLPREYVNAHRLCPHPECPONGSVTCFGEPAQVCACAHY 596
Db      520 CRNYSREGVCTVHCNPLQGEPRFEVHAQCFSCHEPCLPMEGSTONGSGDACRCAHF 579
Qy      597 KDPFCVARGCPSSGKPDLSYMPKFPDEGACQCPINICTSC--VDLDYGCAPAEORA 654
Db      580 RDGPHCVNSCPHGLG--AKGPIYKYPDAQNECRPCHENCTGCGNGPELDQLGOAEVLM 637
Qy      655 SPLTSIVSAVAGILLVVLGVGVGILIKRQOKIR-KYMRRLLOETLEVEPLTSGAMP 713
Db      638 SKPHLVIAVTVG--LAVILMLIGSGLYWRGRIRIQKAMRKRYLERGSEIFELDS-EKA 694
Qy      714 NQAMRLKETELRKVKYLGSAFGVYVYGIIPDGENVKIPEAIKVLRENTSPANKET 773
Db      695 NKYLARLFKEITELRKVKLGSGVGVTHKGIWPEGESIKIPVICIKVIEDKSGOSFOAV 754
Qy      774 LDEAVYVAGVSGPYVRLLGICTSTVOVTOLMPYGCILDHVRNRRGLSGODLLNMC 833
Db      755 TDHMLAVGSLDHAHYRLGLCGSSLOLYTQYLPGLSLDHVKQRETLGPOLLNMGV 814
Qy      834 QIAKMSYLEDELVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETEHADGGKVPK 893
Db      815 QIAKMWYLEEHSMVHRDLARVMMLKSPSQVADFGVADLLPDDKQLHSEAKETIK 874
Qy      894 WMALESLRRRTFHOSDWSYGVTVLMTFGAKPVDGIPAEIPIDLLEKGRLEPQPTC 953
Db      875 WMALESTHFQKYHSDWSYGVTVLMTFGAEPYAGRLAEIDLEKGRRLAQPIC 934
Qy      954 TIDVIMVYKMWIIDECPREFELTSEFSRMAPDQRFVVIQNEIDLGPASPLDSTFYS 1013
Db      935 TIDVYVWYKMWIDINIRTFELANETMARADPRRLVTKRAS-GRGTF-PAAPRS 991
Qy      1014 LLEDDMGDLVDAEELVLPQCGFCFDPAPGAGMVRHRRSSSTRSGGDDITLLEPSE 1073
Db      992 VLTTEL--QEALEPEL-----DLDLLEAE 1017
Qy      1074 E-----EAPRSLAPSEG-----AGSDVPDGDIGMGAKLQ 1105
Db      1018 EGLATLSGALSTPTCTLRPSGQLSPSSGYPMNMOSSIGECALDSAYLAGEQFSR 1077
Qy      1106 SLPTHDPSPLOQYSEDPYVPLDSEITDGYV---APL-----TC-----SPOE---Y 1145
Db      1078 PLSLH-PIPRGR-----PASESSGHVGTGSEAEIQEKVSVCRSRSRSPRPGDSAY 1129
Qy      1146 VNGPDVPRQPPSPREGP-----LPARPAGATLERAKTILSP-GKNGV----- 1187
Db      1130 HSQRHSLTLPTPLSPGLEEDNGYVPPDHLRGASSREGTLSSVGLTGEED 1189
Qy      1188 KDVFARFAGAVENPEYLLTPQGAAPQFHP 1216
Db      1190 ED-----EEYEVNKRKRGRSP-PRP 1209

```

```

OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT      that several genetically defined classes of alleles cluster in
RT      subdomains of the receptor protein.";
RL      Genetics 137:531-550(1994).
RN      [2]
RP      REVISIONS.
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2992499;
RA      Liven E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RT      hormone binding and kinase domains.";
RL      Cell 40:599-607(1985).
RN      [4]
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX      MEDLINE=Oeagon-R; TISSUE=Embryo;
RA      Medline=87002474; PubMed=3093080;
RT      "Alternative 5' exons and tissue-specific expression of the
RT      Drosophila EGF receptor homolog transcripts.";
RL      Cell 46:1091-1101(1996).
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP      ANALYSIS.
RX      MEDLINE=9102120; PubMed=9882502;
RA      Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RT      specification in wild-type, Ellipse, and null mutant Drosophila.";
RL      Dev. Biol. 205:129-144(1999).
RN      [6]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX      STRAIN=Berkeley;
RA      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Branton R.C., Rogers Y.-H.C., Blaise R.G., Champs M., Pfeiffer B.D.,
RA      Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Arif J.F., Abpayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA      Burrows K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA      Foster C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasner K.,
RA      Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegman C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Jiang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclb J.M.,
RA      Palazotto M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svitkeas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195 (2000).
 RN [7]
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekwanryeong;
 RX MEDLINE=85137938; PubMed=2983332;
 RA Wadsworth S.C., Vincent W.S. III, Blodreau-Wentworth D.;
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor";
 RL Nature 314:178-180 (1985).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "Interallelic complementation among DER/fld alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases";
 RL Genetics 129:191-201 (1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor";
 RL Cell 89:13-16 (1997).
 CC - FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOSEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAAS1462.1; -;
 DR EMBL; X03417; AAAS1460.1; -;
 DR EMBL; K03416; AAAS0965.1; -;
 DR EMBL; K03418; AAAS1461.1; -;
 DR EMBL; AF109077; AAD26134.1; -;
 DR EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26131.1; -;
 DR EMBL; AE003454; AAF46732.1; -;
 DR EMBL; X02293; CAA26157.1; -;
 DR EMBL; X78920; CAA55823.1; -;
 DR EMBL; X78919; CAA55821.1; -;
 DR EMBL; X78919; CAA55822.1; -;
 DR PIR; A00640; GORRE.
 DR HSBG; P11562; IFGK.
 DR Flybase; FBgn0003731; Egr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase_1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Trasnemembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 DR Tyrosine-protein kinase; ATP-binding; signal; Alternative splicing;
 DR Developmental protein; 30.
 DR SIGNAL; 1.
 DR CHAIN; 31; 1426; POTENTIAL.
 DR FT; 31; 868; EPIDERMAL GROWTH FACTOR RECEPTOR.
 DR FT; 31; 868; EXTRACELLULAR (POTENTIAL).
 DR TRANSSEM; 869; 889; POTENTIAL.
 DR DOMAIN; 890; 1426; CYTOPLASMIC (POTENTIAL).
 DR FT; 938; 1198; PROTEIN KINASE.
 DR NP BIND; 944; 952; ATP (BY SIMILARITY).
 DR FT; 971; 971; BINDING.
 DR ACT_SITE; 1063; 1063; ATP (BY SIMILARITY).
 DR MOD_RES; 902; 902; PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.2%; Score 1935; DB 1; Length 1426;
 Best Local Similarity 32.4%; Pred. No. 4; 3e-96;
 Matches 464; Conservative 187; Mismatches 435; Indels 344; Gaps 42
 QY 24 QVCTGDMKRLPASPETHIDMLRHLHYGGVQGNLELTYLPT-NASLSPLODIOEVGG 82
 DB 100 KICGTGSRSLSPVSNKHNHRYNRDRYTNCTYVDGMLKTLWLPENULDSFLDNIREVTG 159
 QY 83 YVIAHQVQVPLQRLRIYRGQLP-----ENYVALVLDNDPLNNTPTPGASPGSL 137
 DB 160 YILSHVDVKKVFPKQILIRGTLTSLSVSEKVALFV-----TYSKM 203
 QY 138 RELQRLSLLEILKGVLYLQRPOLCYQDTILWDFHKNQALTLIDTNRACHPSCP 197
 DB 204 YLEIPLRLDVLNGVQGFHNHYVLGMRITQNSEIYNSNTDAVYNDFAIPRECKCHE 263
 QY 198 MCKGSRWQSSSSDQSLRTVYAGGCA--RCKGPLPTDCCHGQCAAGTGPRHSDCLAC 255
 DB 264 SCTHG-CWSGGPKNCKKFSFLTSPQACAGRCVGPAPRCCHLFCAGGCTGPTORDCIAC 322
 QY 256 LHRHSGICELCPALVTYNTDFEPMNPEGRYFGASCYTAACPYNLTSTVGSCITVVC 315
 DB 323 KNFFDEAVSKEBCPRKRYNPTTYLVETNPBEKRYAVGATVACEP-GHLLRNGACVASC 381
 QY 316 PLNQGVTAEDGTQREKSKPCARVCYGLGMNNFTVSWLRVPSASHLSEEPAGCKK 375
 DB 382 PDKMKKGKGF-----CVPGNGCPKTCPG-----VTVLHAGNIDSPRNCITV 422
 QY 376 IFGSLAFLPESFDG--DPSN--TAPQYIKANSK-----FIGITELTGLVLIYSAWDSLPD 427

```

Db 423 IDGNRIIDQTFSSGFODVYANYTWGPRYIPLDPRREVFSTVEITGYLNIESTHPQFN 482
Qy 428 LSVFQNLQVIRGRILHNGAY-SLTLOGISWGLRSLRELSGALIHNTLCTVHTV 486
Db 483 LSYRNLETIHGRQLMSMFALAIVKSYLSLEMRLLKQISSGVVIOHNRDLCTVSMI 542
Qy 487 PMDQLFNRHQAALHTANRPEDECVGEGLAGHQLCARHMGCGFPQCVCSQFLRGQSC 546
Db 543 RMPAIQKEPEKQVWVWNEMLRADLEKNGTICSDQCNDCGKMGATQOCLTKNFNNGTC 602
Qy 547 VEEGRVLOGLFREYVNAHCLPCHPECOQONGSVTCGPADOCVACAHYKDPFCVARC 606
Db 603 IADCGYISNAYK--FDNRCKICHPBCR-----TCMGAGADHCQCEVHRDQHQCVSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PKKYNDRGVCRECHANCDCGTGKDTIGAGCTTCMLAIINDDATVYKGLKDDKCPD- 713
Qy 615 SYMPIWKF--PDEEGACQP-----CPI-----NCTH----- 638
Db 714 GY--FMEXVHPQEOGSLKPLAGRAVCRKCHPLCELCNTNYGHEQVCSKTHYKRQOCET 771
Qy 639 -----SC-----VLDLKG----- 647
Db 772 ECPADHTDEBQRCFOHPECNCGCTPGADDCKSCNFKLPANETGPVYSTMNTCTS 831
Qy 648 -CPAEOR-----ASPLTS-----IVAVGILLVLVGVFGI 679
Db 832 KCPLEMRHVNYQYTAIGPYCAASPPRSSKITANLDVMMIITIGAVLVPITCICLV--T 889
Qy 680 LKRRQOKIRKYT--MRLLQETLVEPLTPSGAMPNOAMRLKETLRKVKVLSGAF 737
Db 890 YICQKQAKKETEYKMTWALSQCDSPKRPINIGALCTLRVYKAEARKGCVLQMGAF 949
Qy 738 GTVYKGIIMPGEVNVKIPVAIKVLRNTSPKANKEIIDEAUVAGVSPYVSLIGICLT 797
Db 950 GRVYKGVWPEGEVNVKIPVAIKELIKETGAESESEFIREVYIASEEHVNLKTLAVCMS 1009
Qy 798 STVOLVTQLMFYGCLDVRNENRGLSQDLLNWCQIAKMSYLEDVRLVHRDLAARV 857
Db 1010 SQMMLITQLMPLGCLLDYVNNRBDKISKALLNWSQIAKMSYLEBEKRLVHRDLAARV 1069
Qy 858 LVKSPNHVKITDEGLARLDIDEVYHADGKVPKIMMALESLRRFTHOSVMSYGT 917
Db 1070 LVQTPSLVKITIDFGLAKLSSDSNEYPAAGGKMPKIMLALECIKNEVFTSKSDVAFVGT 1129
Qy 918 VWELMTFGAKPYDIPAREIPDLLEKGERLPORPCTIIVYVIMVXCMIDSECRPRFE 977
Db 1130 IMWELTFGORPHENIPAKDIPDLIEVGLKLEPEICSLDITCLLSQMLDAMRTPKQ 1189
Qy 978 LVSEFSMARDPORFVVIYQNEDELG--PASPLDSTFYSLELD--DMGDLVDAEYVLP 1032
Db 1190 LTVFAEFARDPGYVLAIPGDKFTFLCA-----YTSODEKLIKRLKLAPTTDSALRK 1242
Qy 1033 QCGFFCPDPAAGAGMWHRHRSSTRSGGDLTLGLEPEEEAP-----RSLAPSRG 1086
Db 1243 PDDYLOPKAAPGS-----HRDCT-----DMPTKLNXYCDDPSKXNS 1281
Qy 1087 AGSDVPDG---DLCMGAKGLQSLPTHDESPLORYSEDPVLPJSTDTGYVALPTCSPOP 1143
Db 1282 TGDDESDSAREVGVGLR-----LDLPVEDEDDYLLMP--TCQPGP 1319
Qy 1144 EYVNPDPVRQOPSPREGLPAARPAQATLERAKTLPKNGVGVAFVFGAGVENPEVL 1203
Db 1320 NNNNNN-----NPNONMAAIVGVAAGVY-----DLIGVPVSDNPEYL 1358
Qy 1204 ----TPGGGAAPOPH-----PPAFSP-AFDNLVYWD 1230
Db 1359 LNAQTIGVGESPIPTQTIGIPVWGQGTMEVKVPMGSEPTSDHEYYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534
DT 21-JUL-1986 (Rel. 01, Created)
DR 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN ERBB
OS Avian leukosis virus.
OC Viruses; Retroviruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crichtenden L.B., Raines M.A., King H.-J.,
RT "c-erbB activation in ALV-induced erythroidleukosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor."
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSSP; P11362; IFKG.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0EB01FCC CRC64;

Query March 25.5; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.4e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

Qy 593 CAHYKPPFCVVARCPGVKPDJSYMPIWKFPEBGAQCPQINCTHSCVLDLDDKCCPAEQ 652
Db 3 CAHFIDGPHCVKACPAGVGENDTL-VWKYADANAVCQLCHNCTRGCKGPGLEGCP--- 58
Qy 653 RASPLSTVSAVY-GILLVWVGVGVGILIKRQOKIRKYTKRRLLOELVELVEPLTPSGA 711
Db 59 NSKTSISIAAGVGVGLLGVVGLGILYLR--HIVRRTLRLLQERELEVEPLTPSGE 117
Qy 712 MPNQAMRLKETLRKVKVLSGAFGVYKGIWIPDGENVAKIPVAIKVLRNTSPKANK 771

```

```
Db 118 APNOAHLRIKETEPKVKVLSGAFGTYYKGLWPEGEKXKIPAIKELREAFSPANK 177
|||
Qy 772 EILDEAYVMAGVSPVSRLLGICLTSTVQVLTQMPYGCILDHVRENRGLSGODLLNW 831
|||
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMPYGCILLDYIREKDNIGQVLLNW 237
|||
Qy 832 CMQIAGKMSYLEEDVRLVHRDLAARVLYKSPHVKITDFGLARLLDIDETFEHADGKVP 891
|||
Db 238 CVQIAGKMWYLEERLVRHDLAARVLYKTPQHVKITDFGLAKLLGADKEXYHAEKGKVP 297
|||
Qy 892 IKMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 951
|||
Db 298 IKMMALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPP 357
|||
Qy 952 ICTIDVYIMVWKCMWIDSECRPRFRELVSFSPMRARDPQRFVVIQ-NEDIGPASPDLSTF 1010
|||
Db 358 ICTIDVYIMVWKCMWIDSECRPRFRELVSFSPMRARDPQRFVVIQ-NEDIGPASPDLSTF 417
|||
Qy 1011 YRSLEDDDMGDLVDAEYLVPOGFFCPDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
|||
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
|||
Qy 1071 PSEEARPSPL-----APSEGASDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVP 1125
|||
Db 450 -----SRTPLLSSAISATSNNSATNCID-----RNGQGHVREDSFVQRYSSDPTGN 495
|||
Qy 1126 LPSET--DGVAPLTCSPOPEVYNOPDVRPOPSPRESPLPAAPAGATLERAKTLSPGK 1183
|||
Db 496 FLEESIDGFL-----PAPEYVNG--LMPKKPS-----TWMQ 526
|||
Qy 1184 NGVVKDVF-----AFCGAVENPEYITPQCGAAPQHPHPAPSPAFDNLVY 1228
|||
Db 527 NGIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPAAKTFESSPY 578
|||
Qy 1229 WQD-----DPPE-----RCAPSTFGKTPAENPEYLGIDVP 1260
|||
Db 579 WTQSGNHQINLDNPYQODFLPNEIKRNGLKVPAENPEYLRVAAP 625
|||
RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein exB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The exB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=8423957; PubMed=6328658;
RA Debutre B., Henry C., Benaisa M., Biserte G., Clavierie J.-M.,
RA Saulie S., Martin P., Stehelin D.;
RT "Sequencing the exB gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -i- Catalytic Activity: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -i- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
```

```
CC -i- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaborati
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on i
CC use by non-profit institutions as long as its content is in no w
CC modified and this statement is not removed. Usage by and for commerci
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; TVYU.
DR PIR; A00644; TVYU.
DR HSSP; P11362; TRFK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM_1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;
Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 14
Qy 593 CAHYKDPFCVACRCPGVKPLDSYPMKPFDEEGACQCPINCTHSCVDLDDKCPAQ 652
|||
Db 3 CAHFDGHCYKACRCPAGVLSGNDLT-VKXADAIVACQLCHPNCRTCKCPGLGECF--- 58
|||
Qy 653 PASPLTGSVSAVY-GILIVVLGVVFGILIKRQOKIRKTYMRLLOETELVEPLTPSGA 711
|||
Db 59 NGSKTPSTAGAVGGLLCTLVVGGIGLYLRR-HYRKRTLRLLOERLVEPLTPSGE 117
|||
Qy 712 MPNOQMRILKETELRKVKVLSGAFGTYYKGLWPEGEKXKIPAIKELREAFSPANK 771
|||
Db 118 APNOAHLRIKETEPKVKVLSGAFGTYYKGLWPEGEKXKIPAIKELREAFSPANK 177
|||
Qy 772 EILDEAYVMAGVSPVSRLLGICLTSTVQVLTQMPYGCILDHVRENRGLSGODLLNW 831
|||
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMPYGCILLDYIREKDNIGQVLLNW 237
|||
Qy 832 CMQIAGKMSYLEEDVRLVHRDLAARVLYKSPHVKITDFGLARLLDIDETFEHADGKVP 891
|||
Db 238 CVQIAGKMWYLEERLVRHDLAARVLYKTPQHVKITDFGLAKLLGADKEXYHAEKGKVP 297
|||
Qy 892 IKMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 951
|||
Db 298 IKMMALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPP 357
|||
Qy 952 ICTIDVYIMVWKCMWIDSECRPRFRELVSFSPMRARDPQRFVVIQ-NEDIGPASPDLSTF 1010
|||
Db 358 ICTIDVYIMVWKCMWIDSECRPRFRELVSFSPMRARDPQRFVVIQ-NEDIGPASPDLSTF 417
|||
Qy 1011 YRSLEDDDMGDLVDAEYLVPOGFFCPDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
|||
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
|||
Qy 1071 PSEEARPSPL-----APSEGASDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVP 1125
|||
```



```

Db      -450 -----SRTLLLSLSTASNNATKCID-----RMGQHPRVREDSFVQRSSDPCTN   495
QY      1126 LBSEET--DGVAVALTSGPPEWYNOPDVRPPCPSPREGPLPAARPGACATLERAKTISPK    1188
                    :|::|||::|
Db      496 FLEESIDDDGL-----PAPEYVNQ-LMPKKPSFM-----                524
QY      1184 NGVAVKVFAF-----GGAVENPEYLTPGGAGAAPQPHPAFAPAD   1224
                    :|::|||::|
Db      525 --VONQNYNRISTATSKLPMOSRYQNOSHSTRAYVDNEL-----NTNQSLACTHYE   574
QY      1225 NLTYWDODPRERGAPPSTFKGPFTLENDEY   1254
                    :|::|||::|
Db      575 SSPYWIQSGNHQ-----INDNDPY   594
RESULT 14
ERBB_AVIEWU STANDARD; PRT:     540 AA.
AC       P11273;
DT       01-JUL-1989 (Rel. 11, Created)
DR       01-JUL-1989 (Rel. 11, Last sequence update)
DI       15-JUN-2002 (Rel. 41, Last annotation update)
DE       Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
CN       v-erbb.
OC       Avian erythroblastosis virus (strain ts167).
OX       Viruses; Retrocid viruseae; Retroviridae; Avian type C retrovirusee.
OX       NCBI_TaxId=103898;
RN       [1]
RP       SEQUENCE FROM N.A.
RX       MEDLINE=87064458; PubMed=2878364;
RA       Choi O.R., Traitor C., Graf T., Beug H., Engel U.D.;
RT       A single amino acid substitution in v-erbB confers a thermolabile
          phenotype to ts167 avian erythroblastosis virus-transformed erythroid
          cells."
RL       Mol. Cell. Biol. 6:1751-1759 (1986).
CC       -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
          tyrosine phosphate.
CC       -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC       CC
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration -
          between the Swiss Institute of Bioinformatics and the EMBL outstation -
          at the European Bioinformatics Institute. There are no restrictions on its
          use by non-profit institutions as long as its content is in no way
          modified and this statement is not removed. Usage by and for commercial
          entities requires a license agreement (see http://www.isb-sib.ch/announce/
          or send an email to licenses@isb-sib.ch).
CC       CC
Cc       EMBL: M13179; AAA42401.1; -.
        DR       PIR: A25231; TVVEBH.
        DR       HSXP_ P11362; LFGR.
        DR       InterPro: IPR000719; Euk_pkinase.
        DR       InterPro: IPR001245; Tyr_pkinase.
        DR       Pfam: PF00069; pkinase; 1.
        DR       ProDom: PD0000001; Euk.pkinase; 1.
        DR       SMART: SMO0219; TYRKc 1.
        DR       PROSITE: PS00107; PROTEIN KINASE ATP. 1.
        DR       PROSITE: PS00109; PROTEIN_KINASE_TYR. 1.
        KM       TRANSFERASE, Tyrosine-protein kinase; ATP-binding; Oncogene;
        KW       Glycoprotein; Phosphorylation.
        FT       DOMAIN              132         399
        FT       NP_BIND             138         146           ATP (BY SIMILARITY).
        FT       BINDING            165         165           ATP (BY SIMILARITY).
        FT       ACT_SITE           257         257           BY SIMILARITY.
        FT       VARIANT            270         270           H->D (IN THERMOLABILE V-ERBB).
        SO       SEQUENCE           540 AA;      60412 MW;      5BS3297AA068B65D CRC64;

Query Match               23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity    55.1%; Pred. No. 2.e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
```

Db	CAHFDGPHCVKAPGAVGSENDTL-VKXYADANNAVQLCHNCRCRGKGGLBSCP----	58
Qy	653 RASPLTISVAAY-GILLVYVIGVGFILIKRROOKIRKYTRMLLOETELVEPLTPSGA	711
Db	59 NGSKTPSIAGVGVGGGLCLVWVGLGIGLTRRR-HIVKRTLRRLQERLEVEPLTPSGE	11
Qy	712 MPNOAKRIKETELRKVKVLGSGAGTGYKGIWPDSENVYIPIAIVLRENTSPYANK	772
Db	118 APNQHRIKETELFKKVKVLGFGAGFTGYKGLMPEGEKTIPIAIELLEAISPANK	177
Qy	772 EILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHYENRGRLGSGDLLNW	831
Db	178 EILDEAYVMASVNPVHCRLGICLTSTVQLITQLMLPGCLLDYREHKDNGISGYLLNW	23
Qy	832 CMQIAKMSVLEDVRLVHRDLAARNVLYSPHNVKITDPCGLARLLDIDETVHADGGVP	891
Db	238 CVQIAKMSVLEDRVHRDLAARNVLYTPQHVKITDPCGLAKQLGADKEKYHAEGGVP	297
Qy	892 IKMAALESILRRFFTHQSDVWSYGYVWELMFGAKPYDGIPIAREIPDLLEKGERLPDP	955
Db	298 IKMAALESILHRIYTHQSDVWSYGYVWELMFGSKPFDGIPIASISVLEKGERLPDP	35
Qy	952 ICTIDVYIMVKKCMMLDESCRPFRFLVSESRMARDQRFVIO-NEDLCPASPLDSTF	107
Db	358 ICTIDVYIMVKKCMMSDADSRPFRELLAEFSKMARDDPRVLYVGDERMLPSPDTSKF	417
Qy	1011 YRSLLIEDDDMGDLVDAEEYLVPOQGFCCPDPAAGAGVHHRHRSSTRSGGDLTGLG	107
Db	418 YRTLMEEDEMDLVDADEVLYVHQGF-----NSPSI-----	444
Qy	1071 PSEEARPSPL-----APSEGAQDVFDGDLGMAKGLQLPLTHDPSPLQRYSEDPTVP	112
Db	450 -----SRPLLSLSLSTATNNSATINCIDNGS-----H-----	476
Qy	1126 LPSETDGVYVPLFCSPQPEYVQDPVRPQPSPREGPIAPAPAGAT-LEBAKTLSPQXN	118
Db	477 -PVREDFGL-----PAPEYVNO-LMPKPESTANVQOITNYSLTAISKLPIDSRYN	527
Qy	1185 GVVYKDVPAFGGAVENTEYL 1203	
Db	528 -----SHSTAVDNPEYL 539	
RESULT 15		
CC	EGFR_CHICK	
CC	ID_EGFR_CHICK	STANDARD; PRT; 703 AA.
AC	P13387;	
DT	01-JAN-1990 (Rel. 13, Created)	
DT	01-JAN-1990 (Rel. 13, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER	
DE	Fragment).	
GN	EGFR.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=88261272; PubMed=3260329;	
RA	Lak I., Johnson A., Howk R., Sapp J., Bellot F., Winkler M.,	
RA	Ullrich A., Vennstrom B., Schlusseringer J., Givol D.:	
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,	
RT	expression in mouse cells, and differential binding of EGF and	
RT	transforming growth factor alpha."	
RL	Mol. Cell. Biol. 8:1970-1978 (1988).	
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.	
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND	
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).	
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein	
CC	tyrosine phosphate.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.	

```

Query Match      22.7%  Score 1558; DB 1; Length 703;
Best Local Similarity 43.3%  Pred. No. 2.7e-76;
Matches 309; Conservative 113; Mismatches 257; Indels 34; Gaps 13,
QY      6 RMGLIALIPGAA-----STQVCTGDMKLRIPASPEHMLDMHLHYGGCGVQGNIE 61
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . .

```

Search completed: July 22, 2003, 08:46:20
Job time : 21.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ-369-383-14

Perfect score: 6853

Sequence: 1 MELALCRMGLLALLPFGA.....TRKPTPTAENPYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 20604715 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phase:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6154	89.8	1259	6	018735
2	3102	45.3	1209	1	09GX70
3	3073	44.8	1210	11	09EP98
4	2688	39.2	1165	13	09YH40
5	2660.5	38.8	1137	13	09G6F6
6	2280	33.3	1138	13	P79754
7	2011.5	29.4	1433	5	09B1H9
8	1871	27.3	419	4	09UX79
9	1739	25.4	367	11	09R2X1
10	1720	25.1	729	15	09E712
11	1718	25.1	567	15	09E714
12	1697.5	24.8	412	4	09WV0
13	1654	24.1	962	5	09A895
14	1645	24.0	545	15	09S468
15	1464.5	21.4	655	11	09WVF5
16	1448.5	21.1	643	11	09ERV6

17	1249	18.2	1193	5	09Y1X8	09Y1X8 ephradya f
18	1188.5	17.3	1368	5	023821	023821 caenorhabdi
19	1179	17.2	1717	5	026566	026566 schistosoma
20	1089	15.9	527	13	090836	090836 gallus gall
21	959.5	14.0	478	11	09S5E0	09S5E0 rattus norv
22	906	13.2	165	4	014256	014256 homo sapien
23	905.5	13.2	599	13	09PSH2	09PSH2 gallus gall
24	887	12.9	176	11	0923V5	0923V5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08S2M1	08S2M1 drosophila
27	754.5	11.0	311	13	09P162	09P162 xiphophorus
28	739.5	10.8	1362	13	09PVZ4	09PVZ4 xenopus lae
29	734	10.7	331	4	09BDU7	09BDU7 homo sapien
30	725	10.6	1671	5	09NJV5	09NJV5 biophalari
31	723	10.6	149	6	09BG66	09BG66 oryctolagus
32	687	10.0	1418	13	093457	093457 scophthalmu
33	683.5	10.0	1368	13	08UM85	08UM85 paralicthy
34	666.5	9.7	1369	13	08UM86	08UM86 paralicthy
35	665	9.7	1422	13	08UM84	08UM84 paralicthy
36	663.5	9.7	1472	5	09U5A8	09U5A8 bombyx mori
37	657	9.6	1358	13	073798	073798 xenopus lae
38	643.5	9.4	1418	13	08UM83	08UM83 paralicthy
39	628	9.2	1245	13	09YGH8	09YGH8 scophthalmu
40	624	9.1	1371	11	09QVW4	09QVW4 rattus sp.
41	618.5	9.0	2144	5	09VD94	09VD94 drosophila
42	570.5	8.6	967	11	09SMR2	09SMR2 mus musculu
43	567.5	8.6	987	11	091YMO	091YMO mus musculu
44	587.5	8.6	1036	4	007912	007912 homo sapien
45	587.5	8.6	1055	11	054967	054967 mus musculu

ALIGNMENTS

RESULT 1
018735 PRELIMINARY; PRT; 1259 AA.
ID 018735
AC 018735;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -;
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L-like; 2.
DR Pfam; PF02757; YLP_2.
DR Prodom; PD000001; Euk kinase; 1.
DR SMART; SM00261; Fur_3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9C4ACD46 CRC64;

Query Match	89.8%; Score 6154; DB 6; Length 1259;	
Best Local Similarity	89.8%; Pred. No. 0;	
Matches 1138; Conservative	48; Mismatches 67; Indels 14; Gaps 4;	
QY	1 MELALCRWGLLLALPPGAASSTVCTGTMKRLPLPASPETHLMLRLHYQGCQVVGNT	60
DB	1 MELAMCRWGLLLALPPGAAGTQVCTGTMKRLPLPASPETHLMLRLHYQGCQVVGNT	60
QY	61 ELTYLPTNASLSFLDIOEVQGVYVIAHQVRYOVLRLRLVRGTOLFEENYALAVLDNG	120
DB	61 ELTYLPTNASLSFLDIOEVQGVYVIAHQVRYOVLRLRLVRGTOLFEENYALAVLDNG	120
QY	121 DPLNNTVTGASPCGLRELOLRSLTEILKGVLIQSNPOLCYQDTILMKDIFHKNNOLA	180
DB	121 DPLNNTVTGASPCGLRELOLRSLTEILKGVLIQSNPOLCYQDTILMKDIFHKNNOLA	180
QY	181 LTLIDTNRBRACHPSPCKGSRGSGESSECCSLTETVAGGARGKGLPTDCCHEQC	240
DB	181 LTLIDTNRBRACHPSPCKGSRGSGESSECCSLTETVAGGARGKGLPTDCCHEQC	240
QY	241 AAGCTGPRGSDCLALFHNHSGICELHCPALVTYNTDTPESMNPBGRYTFGASCTACP	300
DB	241 AAGCTGPRGSDCLALFHNHSGICELHCPALVTYNTDTPESMNPBGRYTFGASCTACP	300
QY	301 YNYLSTDVSGCTIVCPHNCVETAEDGTQCEKSKRCACAVCYGLMNNFTYSFWLR-V-359	359
DB	301 YNYLSTDVSGCTIVCPHNCVETAEDGTQCEKSKRCACAVCYGLMNNFTYSFWLR-V-359	359
QY	360 PKVASHLLEFAGCKKIFGSLAFIPESFDDPASNTPQYIKANSKPIGTETGLYIS	419
DB	360 PKVASHLLEFAGCKKIFGSLAFIPESFDDPASNTPQYIKANSKPIGTETGLYIS	419
QY	420 AMPDLSPLSVFQULQYIRGILHNGAYSLTQGLISWGLRSLBELSGALIHNNTH	479
DB	420 AMPDLSPLSVFQULQYIRGILHNGAYSLTQGLISWGLRSLBELSGALIHNNTH	479
QY	479 LCFVHTVPMDOLFNNPQALHTANRDECEVCGLAHQLCARGCMWGPQCNCSQ	539
DB	479 LCFVHTVPMDOLFNNPQALHTANRDECEVCGLAHQLCARGCMWGPQCNCSQ	539
QY	540 FLRQGEVCEECRVIGQLPREYVNAHCLPCHPECPQNSVTCFGEADQCAVCAHYKDP	599
DB	540 FLRQGEVCEECRVIGQLPREYVNAHCLPCHPECPQNSVTCFGEADQCAVCAHYKDP	599
QY	599 PFCVACPSGVKPDLSMPIMKPFADBEGTQCPINCTHSCALDEKGPASGRASPVS	652
DB	599 PFCVACPSGVKPDLSMPIMKPFADBEGTQCPINCTHSCALDEKGPASGRASPVS	652
QY	660 ILSAVVGLLVVVGAVFGILIKRQOKIRKYMRLLOETELVEBLTSPSGAMPNQAQR	719
DB	660 ILSAVVGLLVVVGAVFGILIKRQOKIRKYMRLLOETELVEBLTSPSGAMPNQAQR	719
QY	712 IIAVVGILLAVVVGAVFGILIKRQOKIRKYMRLLOETELVEBLTSPSGAMPNQAQR	712
DB	712 IIAVVGILLAVVVGAVFGILIKRQOKIRKYMRLLOETELVEBLTSPSGAMPNQAQR	712
QY	720 ILKETELRKVVLGSSGAFVYNGIWPDENVKIPAIKVLRENTSPANKELIDEAV	779
DB	720 ILKETELRKVVLGSSGAFVYNGIWPDENVKIPAIKVLRENTSPANKELIDEAV	779
QY	772 ILKETELRKVVLGSSGAFVYNGIWPDENVKIPAIKVLRENTSPANKELIDEAV	772
DB	772 ILKETELRKVVLGSSGAFVYNGIWPDENVKIPAIKVLRENTSPANKELIDEAV	772
QY	780 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCULDHVRNKGSLGSDLLNMCQIAKAM	839
DB	780 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCULDHVRNKGSLGSDLLNMCQIAKAM	839
QY	832 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCULDHVRNKGSLGSDLLNMCQIAKAM	832
DB	832 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCULDHVRNKGSLGSDLLNMCQIAKAM	832
QY	840 SYLEDVRLVHRDLAARVLYKSRNHVKITDFGLARLLDDETEYHAGGKVPDKMALES	899
DB	840 SYLEDVRLVHRDLAARVLYKSRNHVKITDFGLARLLDDETEYHAGGKVPDKMALES	899
QY	892 SYLEDVRLVHRDLAARVLYKSRNHVKITDFGLARLLDDETEYHAGGKVPDKMALES	892
DB	892 SYLEDVRLVHRDLAARVLYKSRNHVKITDFGLARLLDDETEYHAGGKVPDKMALES	892
QY	900 ILRRFTHQSDVMSYGTVWELMTFGAKPYDGIAREIDLLLEKGRILPOPPICTIDYMM	959
DB	900 ILRRFTHQSDVMSYGTVWELMTFGAKPYDGIAREIDLLLEKGRILPOPPICTIDYMM	959
QY	952 IPRRFTHQSDVMSYGTVWELMTFGAKPYDGIAREIDLLLEKGRILPOPPICTIDYMM	952
DB	952 IPRRFTHQSDVMSYGTVWELMTFGAKPYDGIAREIDLLLEKGRILPOPPICTIDYMM	952
QY	960 IMYCKMVIDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD	1019
DB	960 IMYCKMVIDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD	1019
QY	1012 IMYCKMVIDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD	1012
DB	1012 IMYCKMVIDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD	1012
QY	1020 MGDLDVAEEYLVQQGFPCDPAPRAGAGVNHHRSSSTRSGGDLTLGLEPSEBEARS	1079
DB	1020 MGDLDVAEEYLVQQGFPCDPAPRAGAGVNHHRSSSTRSGGDLTLGLEPSEBEARS	1079

DB	1013 MGDLDVAEEYLVQQGFPCDPAPRAGAGVNHHRSSSTRSGGDLTLGLEPSEBEARS	1079
QY	1080 PLAPSEGAGSDVFDGDLGMGAANKGLSLPTHDSPLOQRYSEDETPVLPSETDVAPLTC	113
DB	1073 PLAPSEGAGSDVFDGDLGMGAANKGLSLPTHDSPLOQRYSEDETPVLPSETDVAPLTC	113
QY	1140 SPOPEYVNPQVPRQPPSPREGPLPAARPAATLER-----AKTISPKGVKXVDFAFG	115
DB	1133 SPOPEYVNPQVPRQPPSPREGPLPAARPAATLER-----AKTISPKGVKXVDFAFG	115
QY	1195 GAVENPEYLTPOGGAAPQPPAFSPAFNLYYMDODPERGAPSTFGTPTAENPEX	125
DB	1193 SAVENPEYLTPOGGAAPQPPAFSPAFNLYYMDODPERGAPSTFGTPTAENPEX	125
QY	1255 LGLDVFPV 1261	
DB	1253 LGLDVFPV 1259	

RESULT 2

Q9QX70	PRELIMINARY; PRT; 1209 AA.
AC	Q9QX70: 01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DR	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Epidermal growth factor receptor.
GN	EGFR.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_Taxid=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RX	MEDLINE=90258886; PubMed=2342466;
RA	Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
RA	Earp H.S.;
RT	"A truncated, secreted form of the epidermal growth factor receptor is
RT	encoded by an alternatively spliced transcript in normal rat tissue.";
RM	Mol. Cell. Biol. 10:2973-2982(1990).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RA	Petch L.A.;
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RA	Gutteridge K.; Dawson T.L.; Earp H.S.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; M37394; AAF14008.1; .
DR	HSSP; P11362; IFGK.
DR	InterPro: IPR000494; EGFR_L domain.
DR	InterPro: IPR000719; Euk_pkinase.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR001245; Tyr_pkinase.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF01030; Recep_L domain; 2.
DR	PRINTS; PR00109; TYRKINASE.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SM00261; Fur; 3.
DR	SMART; SM00219; Tyrc; 1.
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR	PROSITE; PS00107; PROTEIN KINASE DOM; 1.
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SC	SEQUENCE 1209 AA; 134691 MW; 96FE7FECC1B773 CRC64;

Query Match 45.3%; Score 3102; DB 11; Length 1209;
 Best Local Similarity 49.7%; Pred. No. 1.3e-225;

```

Matches :639; Conservative 166; Mismatches 356; Indels 124; Gaps 29;

QY 3 LAALCRMGILLALPPGA-ASTOVCTGDMKRLRPASPTHLDMLRLHYGCGVQGNLE 61
Db 15 LAAICAG-----GALBEKKVCGGTSNRLTQLGTFEDHFLSLQMFNNCEVJGNLE 66
QY 62 LTYLPTNASLSFLDIOEVQGVLIANHQVQLRIYRGTOLEEDNVALAVLNDG 121
Db 67 ITYVQRNYDLSPFKTIQEVAGVLIATLTVRIFLENIQIIRGNALYENTYALAVLSN-- 124
QY 122 PLNNTPTVTSAGPGLRELQRLSLTEILKGVLLQRPOLCYQDTIMKQIFKKNQAL 181
Db 125 -----YGNKIGRLKRLPMNNLOEILIGAVRSPNNPILQNNETIQWRDVI--QDFLSN 175
QY 182 TLIDTRS-RACHPSPMCKGRGCMGSESDCQSLTRTVACGCA-ACKPLPTDCHEQ 239
Db 176 MSMVQRHLTGCPKCDPSCPNRSCWGRGEENCQLTKITICAGQCSRRCRSRSDCHNQ 225
QY 240 CAAGCTGPKHSDCLAHFNHSGICEHCPALVYNTDTESMNPREGRTFGASCYTAC 299
Db 236 CAAGCTGPRSDCLVGRFRDEATCKTCPEPLMYNTTYQMDVNPBGKYSFGATCVKXC 295
QY 300 PYNLTSDVGSCTLVCPPLHQEYTAEDTORCEKSPCARVCYGLGM--FNNFTVSPML 357
Db 296 PRNYVTDHSGCVARACGPDYEV-EDGVSKCKKCDGPRCKVCGIGIGFKD-TLS--- 350
QY 358 RVPKVSASHLEBPAGCKKIFGSLAPLPESFGDPASNTADQVIKANSKFIGITELTGYLY 417
Db 351 ----INATNIGHFKYCTAISGDLHLPLVAFKGDFTPTPLDRELEILKTVKEITGFL 406
QY 418 ISAMPDLPDLVSFQNLQVIRGRILHNGAVSLTLOGIGIMLGIRSLREGSLALIHNN 477
Db 407 IQAMPENMTDLHAEENLEIIRGTRKQHQGSLAVAGNLTSLGRSKIESDSDVITISGN 466
QY 478 THLCFHTVMDQLFRNFHQALLTANRPEDECVEGELACHOLCARGHCWGPPTQVNC 537
Db 467 RNLQYANTINMKKLPFPNCKTKIMNNRAEKDCATNHCNPICSSGCGGPEPTQVCG 526
QY 538 SQFLRGCEVCEVLOGLPREYVNAHCHLPCHECQCPQMGASTVCFPREPDQVCAAHYK 597
Db 527 QNVSRRGECVCKVILBGEPRFEENSECICQHECCIPQMTNITCGRGPDNCKIRKAHAY 586
QY 598 DPFCVACRSPGVPRDLSYMPIMKFPDEGACQPCPINTCHSCVDTDDKCPAEOBASP- 656
Db 587 DGRPCVKTCPSGINGENNTL-VMKFADANNVCHLCHANCTVYCGAGGLKGC--QDEBGRK 643
QY 657 LTIYSAVVGLLVVLCVVFGL-LIKRQCKIRKTYMRPLLOETELVEPLTPSGAMPNQ 715
Db 644 IPSIATGIVGLRIYV-VALGIGLFMRKROLVAKRTLRLLQEREIVELTPSGEAPNQ 702
QY 716 AQMRILKETELRKVKVLSGAFGVYKGIWIPDEENKIPVAIKVIRENTSPANKELID 775
Db 703 AHLRIKETEFKKIKVLSGAFGVYKGLMIPDEKKIKVIAKEIREATSPANKELID 762
QY 776 EAYVMAGVSPYVRRLLIGICTLSVOLVLTQMPYGCILDHVRNRRGLSGODLNNCMOI 835
Db 763 EAYVMASVDPNPHVCRLLIGICTLSVOLVLTQMPYGCILDVREHKNNISQYLNNMCVQI 822
QY 836 AKGMSYLEDVRLVHRDLAARNVLKSPNHRKIDPGLARLLDDEFENYADGKXPITKM 895
Db 823 AKGMNYLEDRLVHRDLAARNVLKTIPOHKITDFGLAKLKGAEKEXYAAEGKVPITKM 882
QY 896 ALBSILRRFTHQSDVMSYGVYWEMLTFGAKPYDGI PAEILPDLLEKERRLPQPICTI 955
Db 883 ALBSILHRIYTHQSDVMSYGVYWEMLTFGSKPYDGI PAEISILLEKERRLPQPICTI 942
QY 956 DVYVIMVCKMNISEGRPRRELVSFSSNARBPQCFVYIQ-NEDIGPSPDLSFTYRSL 1014
Db 943 DVYVIMVCKMNIIDSDSPKRELLILESKAARBPQVLYIQGERHNLSPDLSFTYRSL 1002
QY 1015 LEDDDMGDLVDAEBEYLPQGGFCPPAPGAGGVHRRSSSTRSGGDLTLGJEPSEE 1074
Db 1003 MEEDMEDVDADAEYLLPQGGF-----NSPST----- 1030

QY 1075 EAPRSLAPRSGAGSDVFDGDLGMAKGLQSLPETHPSPLQRYSEDPYPLPSET--DG 11:
Db 1031 --SRTPLLSSLSANSN-----SSTVACTINRNGSCVKKEDAFLOQYSDDPTVLTEDNIDPT 106
QY 1133 YVAPLTCSPOEYVNPDPVFPQPSPEPGLPAAPAPGATLERAKTSLPKQNGVXDVFA 11:
Db 1085 FL-----FVPEYINQ-SVPRKPRAGSQNPVYHNQPLHP-----APGRDLHYQN--P 11:
QY 1193 FGAVANENYI--TQCGAAPQHPHPFPSPAFDNLYWDO-----DP-----PE 12:
Db 1128 HSNMANSPEYINTAQ-----FTLSSGFSFSSALMIGKSHQMSLDNPDYQODFFPK 11:
QY 1236 RGAPSTFKTPTAENPEYICLDVP 1260
Db 1179 EAKPNGIFKG-PTAENAEYLRVAP 1202

RESULT 3
Q9EP98
AC Q9EP98 PRELIMINARY; PRT; 1210 AA.
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Singhal C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahtle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Singhal C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahtle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAC28045.1; -.
DR EMBL; AF275364; AAC28045.1; JOINED.
DR EMBL; AF275365; AAC28045.1; JOINED.
DR EMBL; AF275367; AAC24386.1; -.
DR HSSP; P11362; IFGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; Egr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_cmr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU_5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.

```

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TIR_1.
 DR ATP-binding; Receptor; transferase.
 SQ AQP-binding; Receptor; transferase.
 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.8%; Score 3073; DB 11; Length 1210;
 Best Local Similarity 49.2%; Pred No. 26-223;
 Matches 630; Conservative 168; Mismatches 362; Indels 120; Gaps 26;

QY 11 LLLALLPGAA--STQVCTGTDMKLELPASPETHLMDLRHLYOGGQVVGNNLELTPTN 68
 14 LLLALCAAGALEEKVCOGTSNRLTQGTGFEDHLSLQRMNCEVYLGNEITFYQRN 73
 QY 69 ASLSLQDIQVQVYVLAHQVQVPLQRLIYVGTGLFEDNALATLDNGDPLNNTTP 128
 74 YDLSFLKTIQEVAGVYVLAHQVQVPLQRLIYVGTGLFEDNALATLDNGDPLNNTTP 124
 QY 129 VTGASPGQLREQLRSLEILKGVYVLAHQVQVPLQRLIYVGTGLFEDNALATLDNGDPLNNTTP 184
 125 -YGTNRGTRELPMRLQETIIGAVRFNNPILCMWDITQWRDIQVNVFMSNMDL--- 180
 QY 185 DTRSRACHPCSPMCKSGRCWSESSDQSLRTYVAGGCA-RCKGPIPTCCGECQAG 243
 181 -QSHPSCKCPDSCPCNSGSCWGGGECNCKLTKIICAQCCSHRCGRSPDCCCHNQAG 239
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFEEMPRNEGRTYFGASCYACPYNY 303
 240 CTGPRRESCLVCGCKPDDEATCKDTPRLMLNPTTYQDVNPBGKYSFGATCVKCKRNY 299
 QY 304 LSTDVGCTLVCPRLNQEYTAEDGTQRCCKSKPCARVYCYGLM--FNNFTVSFWLBPVK 361
 300 VYTDHGSQVRAQGPVYEV--EDGQTRKCKCDGPRKQCNIGIGEFKQ--TLS----- 350
 QY 362 VSASHLEEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSFKIGTETGLYISAW 421
 351 INNTNKHFKYCTAISGLDHLIPVAFKDDSPRTPLDRLRELILKTYKELTGLLIQAW 410
 QY 422 PSLPDLVFNQLQVIRGRILHNGAYSLTQGLISWGLSLRLSGALILHNHNLIC 481
 411 PDMWTLHAFEMLEIRGRTHQHGQFSLAVVGLNITSLGLSLKEISGDVITISGNNLIC 470
 QY 482 FHTVPMQDLFRNPQALHTANREDECYEGGLACHQLCARGCMGGRPCNCGQFL 541
 471 YANTIMKKLFGTQMKTKYNNRBEKCKAVNHCNPLCSSGCGWGEPRDVCSCQNV 530
 QY 542 RQECVCECVLQGLPREYVNAHGLPCHPRECPONGSVTCFGEADQCVACAYKDPF 601
 531 RRECEVEKCNILEGPREFEVENSECICQHPCLPQAMNITCTGSGPDMCICAHYIDGP 590
 QY 602 CVARCSGVKPLSLYMPKFPDEGACQPCPINTGSCVLDKXGPAERASPLTSIV 661
 591 CVKTCFAGIMGENNTL-VKTVADANNVCHLGHANTYACAGPGLQGEVWPSGKPSIA 649
 QY 662 SAVVGLLVVVGAVFGI-LIKRQCKIRKTYMRLLOETELVLEPLTPSGAMPQAOMRI 720
 650 TGIYGLPLFIV-VALGIGLFMRHRHRYRKRTLRLLQERLVEPLTPSGAMPQAOALRI 708
 QY 721 LKTELAKYKVLGSGAFVYKGIIPGGENVKIPVAIKVRENTSPKANKEIDEAYVM 780
 709 LKTEFEKKIKVIGSAGFVYKGLMIPGEKVKIPVAIKVRENTSPKANKEIDEAYVM 768
 QY 781 AGVSPVYSRLIGICTSTVQVLTQMPYGCILDPHRENRGLASQDLMMQOIAKMS 840
 769 ASVDNPHVCRLLIGICTSTVQVLTQMPYGCILDPHRENRGLASQDLMMQOIAKMS 828
 QY 841 YLEDVRLVHRDLAANVLKSPNHVKITDFGLARLLDIDETEVADGKYPKIMMALESI 900
 829 YLEDRLVHRDLAANVLKTPQHYKIDFGLAKLGAEEKEXYHAEGKVPKIMMALESI 888
 QY 901 LRRFTHSDVMSYGVYVWELMTFGAKYDDGIPAREIPDLLEKERRLPPEPCTIDVYMI 960
 889 LRRFTHSDVMSYGVYVWELMTFGAKYDDGIPAREIPDLLEKERRLPPEPCTIDVYMI 940

DB 889 LRRFTHSDVMSYGVYVWELMTFGAKYDDGIPASDISLLEKERRLPPEPCTIDVYMI 940
 QY 961 MYKCMWIDSECRPRELIVSEFSMAARDPQRFVYIQ-NEDLGASPLDSTFYRSLEDD 10:
 DB 949 MYKCMWIDSECRPRELIVSEFSMAARDPQRFVYIQ-NEDLGASPLDSTFYRSLEDD 10:
 QY 1020 MGDVDAEYVLPQGFPCPDPAQAGAMVHRRSSSTRSGGDDLGLGEPSEEARPS 10:
 DB 1009 MEDVDADAEYVLPQGFPCPDPAQAGAMVHRRSSSTRSGGDDLGLGEPSEEARPS 10:
 QY 1080 PLAPSEGASDVFDGDLGMAKGLQSLPTHPSPRLQRYSEDPTVPYPSER--DGYVAPL 11:
 DB 1035 PLSSISATSN---NSTVACINRGSCRYKEDAFLOQYSSDPTGAVTEDNIDDAFL-- 10:
 QY 1138 TCSPPQEVYNQDVPQPSPREGLPLAPRAGATLEBAKTLSPGKGVYDVAFGAV 11:
 DB 1088 ---PVEYVNO-SVPRKAGSVQNVYHNOPLH-----AGRDLHYON--PHSNV 11:
 QY 1198 ENPEYL-TPQGAAPQPPAPFSPAFNLYWDO-----DP-----PERGAP 12:
 DB 1134 GNPEYLNTAQ-----PTCLSGFRSPALMTQKSHQMSLDNPYQODFFKETKPN 11:
 QY 1241 STFGKTPTANPEYGLDVP 1260
 DB 1185 GIFKG-PTAENAEYLRLVAP 1203

RESULT 4

QY940 PRELIMINARY; PRT; 1165 AA.
 AC QY940;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_Taxid=6086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 overexpression and mutational alterations.",
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Schartl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53471; AAD10500.2; --
 DR HSSP: P11362; IFGK.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_Kinase; 1.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: PS00107; PROTEIN KINASE_ATP_1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM_2.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR_1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 757EB3BD8771A74E CRC64;

Query Match 39.2%; Score 2688; DB 13; Length 1165;
 Best Local Similarity 45.0%; Pred. No. 2,66-194;
 Matches 575; Conservative 172; Mismatches 384; Indels 148; Gaps 31;

QY 1 MELALCRWGLLMLPRG-AAST-----QVCTGDMRLRLPASETHLMDLRLYLGGCV 55
 DB 4 LLELEL-----LLELLLSIGRCCSTDPKRCQGSNQM--LDNLYLKKKKMYSGCV 56
 QY 56 VQNLLETLYPTNLSLFLQDIQEVGVYLIANHQVQVPLQRLRIYRGTQLFEDYVALA 115
 DB 57 VLENLEITYQENDLSFLQSIQEVGVYLIANNEVETIPLVNLRLRGNLKEGFTLL 116
 QY 116 VLDNDPLANTTPTVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYDTILMKDIFRK 175
 DB 117 VMSNYQR-NPSSP--DYVQVGLKQLQLSNLTLEILSGVKNHNPFLCNVETIMMDIVDK 173
 QY 176 NNQALATLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRITVCAGGC-ARCKGFLPTD 234
 DB 174 TSNPTNMLIPIHAFERQCKDPCGVNCSWAAPRGHCQKFTKLCAEQCNRRRGKPID 233
 QY 235 CCEHQAAGCTGPRHSDCLACLAHFNHSGICELHCPALVTYNTDTFESMNPDEGRYFGAS 294
 DB 234 CCNEHAGAGCTGPRATDCLACRDFNDGTCKDTPPKIYIVSHQVVDNPIKITYFGAA 293
 QY 295 CVTACPRVYLTSDVGSCTLVCPRLNCEVTAEDGQREKSKCARCYCLGMFN-NFTV 353
 DB 294 CKECPNSVYVTE-GACVRSASAMLEVD-ENGRSKPCDGVCPKVCDDIGISLSNTI 351
 QY 354 SFWLRVPRVSASHLEEFAGCKKIFGSLAFPESEFGDPASNTAPQYIKANSKFIQTEYT 413
 DB 352 A-----VNSTNGSFENCTKINGDIILNNSPEGPHYKIGMPDBEHLMTYKELT 404
 QY 414 GYLIYAMPDLSPLSLVFNQVQVIRGLINNGAVS-LTLOGLSIMLGLSLBELSGSLA 472
 DB 405 GYLVIMMPENNKTSLSVFQNLRIIRGRTFSRGSPFVVVSHQVLSLLEVSAGNV 464
 QY 473 LIHNHTLCFVHTVPMQDLFRNPQALLHTANRPEDECEVGEGLACHQLCARHCMPGPT 532
 DB 465 ILKNTPOLRVASTITMNRRLFRSEDSQSLSEYDART-----ENQTCNNEGSEDCWPGPT 517
 QY 533 QCVNCSQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECQPNQNSVTFCFEADQCYA 592
 DB 518 MCVSCLHYDRGRCVASCNLIQGEPREAOVDGRVCVCHQCECLVQTDLSLCYGGPANCSK 577
 QY 593 CAHYKDPFCVARGCPGKPDLSYMPMKPDEDEGAQCPRLNCTHSQVDLDKXGPAQ 652
 DB 578 CAHFQDGPQCIPTPRPHGLDGDITL-IMKADKKAGQCPQHQNCQCGGPGISGGRG- 635
 QY 653 RASPLTSTVSAVAVGILLVVLGVVFGILLIKRRQOKIRKTYMRLLQETELVEELTPSGAM 712
 DB 636 IVSHSSLAVGLVSGLLITVIALIIVLLRRRIK-RKRTIRLLQEKELVEELTPSGA 694
 QY 713 PNOQMRLIKETELRKXKULGSGAGFTVYVYGMIPGCENKIIVALKVIRENTSPANKE 772
 DB 695 PNOAFLLIKETELRKORVLSGSGFTVYVYGLMMPDENRIRIVALKVIRENTSPANKE 754
 QY 773 ILDEAYVAGVGSFVYVRLGICLTSTVQLVTLQMLPEYGLLDHVRNRRGLSGQDILANC 832
 DB 755 VLDEAYVAVSYDHHVCRLLGICLTSAVQLVTLQMLPEYGLLDVRRQHRIGQMLANC 814
 QY 833 MQIAKMSYLEVDVLLVNRDLAARVLYKSPNHYKIDFGIARLLIDDEFYHADGSKVPI 892
 DB 815 VOIAKGNVLEEERLLVARDLAARVLLKKNHVKITIDFGISKLTLADEKEYQAHGSKVPI 874
 QY 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDLEKGRPLPQPI 952
 DB 875 KMALESILQMTYHQSDVMSYGVTVWELMTFGSKPYDGIIPAKELASVLENERLPPQPI 934

QY 953 CTIDVYMLMYKCMIDSECRPRELVSFPRMAADPQRFVIONEDLGPASPLDSTFYR 101
 DB 935 CTIEVYMLILKCMIDSSRRFRFELVGEFSQMARDSRLVLVQ--NLPSSDRRLFS 991
 QY 1013 SLEDDMGDLVDAEYLYVPOGFFCPDPAPGAGVYHRRSSSTRSGGDLTLGLEPS 107
 DB 992 RLSSDD--DVYDADEYLL-----RYKIN-RQGS----- 101
 QY 1073 EEEARPSLASEAGSDVFDGLGMAAGLQSLPHDSPLQORSSEDTY-PLRSETD 113
 DB 1019 -----EPCIPPNH-----PVRENSIALRTISDPTQNALEXDLD 105
 QY 1132 GYVAPLTCSPQPEYVNPDPVROP-----PSPRE-----GPLP-AAPAGATLERAKT 117
 DB 1053 GH-----EYVNPQSGSETSRSLSDINPNYEDLTQGWGVSLSSQAEATNFSRPEY 110
 QY 1179 LSPKNGVYKDVFAFGAVENPEYLYTFQGAADQPHPPAFSFAFDNLVYWDODPEPGA 123
 DB 1103 LNTQNSL---PLVSSGSMDDPDY---QAQ-----YQAAF-----LPOTGA 113
 QY 1239 PSTFTKGPPTAENPEYLG 1257
 DB 1138 LTGNGMFLPAENLEYGL 1156

RESULT 5

Q9MEF6 PRELIMINARY; PRT; 1137 AA.

AC Q9MEF6
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERB4.
 OS Gallus gallus (chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 X MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nrg1) and erb4 transcripts in
 RT embryonic chick hindbrain."
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR EMBL: AF121963; AAD31764.1; -
 DR HSSP: P11362.1FGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001368; TNFR_c6.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Dkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 1.
 DR Pfam: PF02757; YLP_2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SMO0261; FU_3.
 DR SMART: SMO0219; TYKc.1.
 DR PROSITE: PS00107; PROTEIN KINASE_ATP_1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM_1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR_1.
 DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

38.8%; Score 2660.5; DB 13; Length 1137;

Best Local Similarity 46.4%; Pred. No. 3,1e-192;
Matches 531; Conservative 171; Mismatches 360; Indels 83; Gaps 25;

```

QY 161 LCYQDTILMKDI FHKNNQLALTITDNRSRACHPCSPCKSCRCWGESSEDCQGLTRTV
DB 3 LCFADITHMODIVRNPMASNFTLVPTNSSGCGRCHKSCCTG-RCWGPTEHNCQTLTCTVC 61
QY 221 AGGC-ARCGPLPTDCHCQCAAGCTGPRHSDCLAHLPHNSGICELJHCPALVYNTDTF 279
DB 62 AEGCDGRCTGPIVSDCHRCAGCGSPDPTDCCAFKMFNDGACVTCQPGTFYVNPPTTF 121
QY 280 ESMNPDEGRYTFGASCTACAPYNVLTSDVGSCTLVCPLNQEVTAEDGTQCEKSKPCA 339
DB 122 QLEHNNAKYTYGAFCKKCPHNFV-VDSSCVCRAPSSKMEV-EENGIKKCKCTDICTP 179
QY 340 RVCYGLGEMNFTVSEFWLRVPRKVSASHLEEFAGCKKIGSLAFIPESDGPASNTAPQY 399
DB 180 KACDGI-----TSLVSAQTVSDSNIDKFINCKTNGNLIPLVTGIGHGDPYHTIAIN 233
QY 400 IKANSKFIGITELTGLYLSAMPDSLPLSVFQNLQVIRGRILHNGAVSLTQGLGISWL 459
DB 234 PEKLNIFQTVREITGLYNIQSWPEKMTDFRVSNSLVITIGRALVSGLSLILKQGITSL 293
QY 460 GLRLRELGSLALHNHNLHCFVTVWMDLFRNPHALHTANRPEDCVGEGLAHQ 519
DB 294 QFOSLKQISAGNIVTIDNSNLCTVHTVMTSLFSTPSQKTIHKKKAENCTADGMYCNE 353
QY 520 LCARGHMGPRPTQCNCSQPLRQGECEVECRVLQGLPREVYNAHCLPCHPECP-ONG 578
DB 354 LCSDDCMGPRDQDLSCRFIRGCTIESCNLYGFERFRANSVCMECDPOCKEKEDN 413
QY 579 SVTCFGEPEADQCAAHYKDPFCAPCPGVKPDLSYMPIKFPDEGACQPCPINC 638
DB 414 MTCYGPGRDHCTCKCFHFQDGNCEKCPDGLQGANSE--FFKXADEDRCHCPHNCQ 471
QY 639 SCVDLDDKGC-----PAQRASPLTISYSAV-GLLVVYGVVGLIKRQ 666
DB 472 GGRGASHDCIYPTPTROSTLPQNR-TPL-IAIGVIGGFLITVINGLTFAYVVRKST 528
QY 687 KIRKYTMRRLQETELVEPLTPSGAMPQAMRILKETELRKVVLGSGAGTVYKGIW 746
DB 529 K-KKRALRFL-ETELVEPLTPSGTAPQADRLIKETELRKVVLGSGAGTVYKGIW 586
QY 747 PDGENVKIPVAKVLRNTSPKANKELIDEAVMAGVSPVSNLGLCLTSTQVLTOL 806
DB 587 PGEETVKLPVAKILNETTPGPANVEFDEALIMASMDPHVLVLGVCLSPITQVLTOL 646
QY 807 MPYGLLDHVENRGLSGODLLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPHYK 866
DB 647 MHPGCLDHYEHKONIGSOLLNMCQIAKMYLEERRLVHRDLAARVLYKSPHYK 706
QY 867 ITPDGLARLLDIDEHYADGKVPKIMMALESLIRRAFTHQSVNSYGVTVWELMTGA 926
DB 707 ITPDGLARLLEDGEVYADGKVPKIMMALECHYKFTHQSDVMSYGVTVWELMTFGG 766
QY 927 KPYDGIPIAREIPLDLKGERLPQPICTIDVYIMVCMWIDSECRPFREIYSEFSRMA 986
DB 767 KEYDGIPTREIPDLKGERLPQPICTIDVYIMVCMWIDASRKPFELAAEFSRMA 826
QY 987 RDPQRFVITQNE-D-LGPASPLDSTYRSLLEDMDMDVIDAEVLLVQDGFPCPDPAFGA 1045
DB 827 RDPQRFVITQDGRKLPSPNDSKFFQTLDEEDLEMDMDAEVLLV-QAFNIPPTLYTS 885
QY 1046 GGMVHRHRRSSSTRSGGDLTLGLEPSESEAPRS--PLAP-SEGAGSDVDGDLGMAAK 1102
DB 886 RTRIDSNRNOFYRRGGYAAEGV-PMYRARPGLIIEBAPVQAATIEFEDCCNGTLR 944
QY 1103 GLQSLPTHDPBLQRYSEDPTVPLPS-----ETDGVAPLTCSPPQEVYNGPDVAPQ 1155
DB 945 KQAVTAKEDSSTORYSADPTVFIPEKVIKGELEDDEGMYTMRKPKXTDYLANVEENPFV 1004
QY 1156 PSPREGPLPAA-RPAGATLERAKTLSPKNGVVDV-----AGGAVENPEVLTITQ 1206

```

```

DB 1005 SRKKXGDLQAVDNPPEYHN-----APNGCPKADDEVYNEPLVINTFANTLENAYL--- 107
QY 1207 GGAPDPHPPPAPFSAFDNLYWQDPPERGA--PSTFKGTPT-----AE 121
DB 1055 -----KNLPEKAKKAFNDPDMYHNSLPPRSTLQHPDVLQYSTKYFGKNGRIRPIAE 114
QY 1251 NPEYL 1255
DB 1110 NPEYL 1114

```

RESULT 6

```

P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TRENBLREL. 03, Created)
DT 01-MAY-1997 (TRENBLREL. 03, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Erib3.
US Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_Taxid=31033;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=99177477; PubMed=10077531;
RA Gellner K., Brenner S.,
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -
DR HSP; P11362; 1FGK
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATF; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

```

Query Match

Best Local Similarity 33.3%; Score 2280; DB 13; Length 1328;
Matches 518; Conservative 153; Mismatches 425; Indels 196; Gaps 31

```

QY 9 WGLTALALPP--GAATQ---VCTGDMKLRLPASPETHLMLRHLHYOCQVQVQNLTL 62
DB 4 WRLIMCYASLALRAASQGTQAVPGTQNGLSSTGSEONYNLNRKRYKGCETIMGLET 63
QY 63 TYLPTNASLSLFDIIGEVGYVLIHNVQVPLQRLRIYRGTQLFEEDYNAVALDNGDP 122
DB 64 TOIESNMDFSLKTIKREVTGVLIAHMHFQEIPLQGLRVRIGNSLVERRALSVFLN--- 120
QY 123 LNNTPVTVGAPGGLRELOQRLTELKGGVLLQRPNLQCYOYOTILMKDIFHKNNQLALT 162
DB 121 ----YPKDG--PSGLNQGLMNLLEIIDGAVQIINNYLRYEGVYWRDIT-RNNDAPLE 173
QY 183 LIDNRSRACHPCSPMKGRCWGESSEDCQSLTRVCAAGC-ARCGPLPTDCHCQCA 241
DB 174 IOFNGERGVC---KSC-GNYCWGPGKQDQQLITKVCAPQCNDRCFGSPRCCIHIEA 229
QY 242 AGCGPKHSDCLALHNNHSGICELHRLPALVYNTDTFESMPDEGRYTFGASCTVAC 301
DB 230 AGCGPLDPTDCAFRLPNDGACVCPQCPQTLINQKQTFQMETPNKAYGYGSIQSCPT 289

```



```

QY 302 NYLSTDVSGCTLVCPHLNQEVTAEDGTOR-CEKSKSCPCARVCYGLGMFNNFTVSWMLRV 360
DB 290 HFV-VDSSSVSVCPDMEV--ERGSQRCELCISGCPKCESTGAEQGT----- 338
QY 361 KVSASHLEEFACCKIFPSLAFLESPFSDPASTAPQYIKANSKFITELGYLYISA 420
DB 339 -VDSSNIDSFNCKTIQSLHFLTGLGDDPKVVPPLDAKKEVFRTVEIDIDINIOS 397
QY 421 WPDSPDLVSFQNLQVIRGRILHNGAYSLTQGLGIMWLGASLREGGSGALIHNTHL 480
DB 398 WPKELNDLSVFSLSLTIGRSLFRKFSLMWYRIPTLISGLRSIRESDSVYISQNAHL 457
QY 481 CFVTVPMWDLFRNH-QALLHTANRPDESCVGEGLACHOLCARGHGCGPPTQVNC 539
DB 458 CYHHTVNTWTLFRGSRVANSINSRPMACEVADGRVCDPLCSGCGWGPQCLSCRN 517
QY 540 FLRGOECVEECRVLQGLPREYVNAH-GLRPHPCOPONGSVTCFGEADOCVACAHYD 598
DB 518 YSRHGTCAAGCHFNHSGIPREFAGLNGCVACHPECKFOTGKASTGTGADECHAKTKFRD 577
QY 599 PEFVACRCPGKPDLSYMPIWKEPDEEGACQPCPINCSTHSCVDLDDKGPASQAPL 658
DB 578 GPYCMSSCPAGVN-DGEKGLIFKFPNREGHCEPHQNCOTQCGSGPLNDC--LEAARLT 633
QY 659 SIVSAVVGILLVVLGVF-----GILLRQOKIRKXTMRLLDETLEVLPLPSGA 711
DB 634 ISSGQITGIALGVPAGLFCLVLFELGMLYHKGALIRKAMRYLESGSEFPPLP-GE 692
QY 712 MPMQAKRILKETELRKVVLGSAFGTVYKGIWIPGENVKIPVAIKVIRENTSPANK 771
DB 693 KGTXYHARILKPSDLRIKPLGSGVFGTVSKGPMIPGEYTKIPVALIKTIQDSSGQGT 752
QY 772 EILDAYMAGVGSYYSRLIGICTSTVQLVTLQMLYGCCLDHVRENRGLSGQDLNW 831
DB 753 EITDHLMSGSLDHPYIVRLIGICPGCLQVLTQSSHGLLEIRHOKTSLPQRLNW 812
QY 832 CMOIAKMSYLEDEVLRHDLAARNVLVKSNNHKTDPGLARLDIDEHYADGKVP 891
DB 813 CVOIAKMYILEHRVYHKLAAKNILKNDYQCIDYVADLLYDDCKVYSETKTP 872
QY 892 IKMMALESILRRRTFTHSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLKGERL 951
DB 873 IKMMALESILRRRTFTHSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLKGERL 932
QY 952 ICTIDVYIMYKCMWIDSECRPRELVSERSPMARPOCFVITQNEDDLGPASPLDSTY 1011
DB 933 ICTIDVYIMYKCMWIDSECRPRELVSERSPMARPOCFVITQNEDDLGPASPLDSTY 980
QY 1012 RSLLEDDMDGLVDAEYLYVPOGFCPPDAPAGAGVWHHRSSSTRSGGDLTLGLEP 1071
DB 981 ----EDSGMGFL-----RGSER--GILLEADLEB 1004
QY 1072 SEEPARFRPLAPSEGAGSDVFDGLMG--AAKGLOSPTHDPSPLO-----RY 1118
DB 1005 DEEE-----GLGDRFATPSLOPSPSWSTSPQINSYVWMTQLRY 1043
QY 1119 SEDPTVPLPSETDGVAPLTCSPOP-EYVNO-----PVYRQPSPREGL 1163
DB 1044 D-----FAVSOGGHIGYLPMSBPVITIQMLYQORSLSSVATLPRRAFRRSSSEAE 1097
QY 1164 --PAARPAAGATLEAKTSLPEKNGVXKDVAFAGAVENPEYLLTPQGAAPQHPAPAFSP 1221
DB 1098 CEDGAOCAGIFRVE-----FGSERGN-----POGG----- 1122
QY 1222 AFDNLVYWDODPPERGAPSPSTFKGTFAENPE 1253
DB 1123 -----QQRKSLSTASPSFSFKTMADDEDE 1146

```

RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles
OX NCBI_Taxid:7165;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SU4.
RA "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor".
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000345; CyC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00699; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
FT Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 29.4%; Score 2011.5; DB:5; Length 1433;
Beet Local Similarity 32.3%; Pred.No.57e-143;
Matches 469; Conservative 197; Mismatches 396; Indels 391; Gaps 37
QY 26 CTGDMKRLRPLASPETHLDMRLHLYQSCQVYQGNLELTYPTVASISFLDIOEGVGYL 85
DB 1 CIGTNGRMSVAPNREHYHKNLRDXYTCTVDGNGLEITWQNTDNLFLQIHRELVGYL 60
QY 86 IAHQVQVPLQRIKRYRGTLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
DB 61 ISLYDLPOVILPRQIIRGRITFTFLNKMEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGVILIQNPQCYQDITLWMDI-FHNKNOLATLIDNRSRACHPSPMC 199
DB 105 ELPLRLDILGSGVFPNNYVNLCHKXSNMBEILLAPOTSQYTFNFSRPRVCPCHPSC 164
QY 200 KGRSCWGESSEDCSLTRVYCAAGCA--RCKGPLPTDCCHEQCAAGCTGPKSHDCLACH 257
DB 165 EVG-CWGEAGANCRCFSKLNCSPQSGRCFGKRPKECHLFCAGGCTGTFGSDCLACKN 223
QY 258 FNHSGICELHCPRLVTVNTDTEFESMPREERYTFGASCVTACGYNTLSDVSGCTIVCL 317
DB 224 FYDGVCKQPCRPQIINPTNYFMERPPDQKAYAGTCVAKCP-EHLKDNQACVAKCPK 282
QY 318 HNOETVADGTORCEKSKPCARVCYGLGMFNNFTVSWMLRVKVSASHLEEFACCKIF 377
DB 283 GRMPQNSE-----CVPCKGVCPKTCGEGL-----VASDNYGNKQCTIIE 323
QY 378 GSIAFLPESFDGPASTN---APQYIKAN-----SKFTGITELTGYLYISANPDSLPDIS 429
DB 324 GSLEILDQSPDGFQOQVYTNFSFSPRYIKIDIPRLLEVSVTKEITGFINIQAHNPFTLIN 383

```


RN (1)
 RA SEQUENCE FROM N.A.
 RP Strausberg R.;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC027080; AAH27080.1;
 SK Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E10180 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 3,1e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESLIRRTTHOSDWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKEERLPDPPICT 954
 Db 1 MALESILRRRTTHOSDWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKEERLPDPPICT 60
 QY 955 IDVYIMWKCMWIDSECRPRFRELVSERPMARPCQSFVVIQNEEDLGPSPMDSTYRSI 1014
 Db 61 IDVYIMWKCMWIDSECRPRFRELVSERPMARPCQSFVVIQNEEDLGPSPMDSTYRSI 120
 QY 1015 LEDDDMGDLVAEEYLVPOGFCPPDPAFGAGMVRHRRSSSTRSGGDLTLGLEPSEE 1074
 Db 121 LEDDMGELVAEEYLVPOGFCPPDPAFGAGMVRHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPLAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPYPLPSETDGY 1134
 Db 181 EPPRSPLAPSEAGSDVFDGDLAVGVTKGLQSLPHDLSPLQYSEDPYPLPSETDGY 240
 QY 1135 APLTCSPOPEYVNPQDVAPPPSPREBPPLPAAREAGATLERAKTLSPGKGVKDVFAFG 1194
 Db 241 APLACSPQPEYVNPQDVAPPPSPREBPPLPAAREAGATLERAKTLSPGKGVKDVFAFG 300
 QY 1195 GAVENPEYLTPOGGAAPQHPFPAFSPAFDNLVYWDODPPERGAPSTFGITTAENPEY 1254
 Db 301 GAVENPEYLTAPAGTASQHPFPAFSPAFDNLVYWDQNSEQGPSTFGITTAENPEY 360
 QY 1255 LGLDVPV 1261
 Db 361 LGLDVPV 367
 RESULT 10
 Q086712 PRELIMINARY; PRT; 729 AA.
 AC Q086712;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 CX NCB1_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennistrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60725.1; -;
 DR HSSP: P03322; 1A6S.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D26914EF1D53 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 2,4e-121;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTCFGEADQCVACHYKDPRECVARCPBGVYFPLDSTYPMWIFPPEEGACQCPPI 634
 Db 141 PEETATPPTGP--DHCKCAFFIDGFCVACAPGAVGENDTL--VMKYADANAVCQLCHP 197
 QY 635 NCTHSCVDLDKCGCPACORASPLTSIYSAVY-GILLVVLGVVFGILIKRQOKIRKRYTM 693
 Db 198 NCTGCKCPGLEGGP---NSKTPSIAGVAVGGLCLVVGVLGILYLR--HIVKRTI 253
 QY 694 RRLIQTETLVEPLPLPSGAMPNQAOMRLKTELRKYVLDGSAFGYVYKIMIPDGENVK 753
 Db 254 RRLIQERLVEPLPLPSGAPNQAOLRLKTELRKYVLDGSAFGYVYKIMIPDGENVK 313
 QY 754 IPVAIKVLRNTSPKAKKEILDEAYVWAGVSPVSRLLGICLSTVQVLTOLMPYGCIL 813
 Db 314 IPVAIKELRENTSPKAKKEILDEAYVWASVDNPRVCLLGLCLSTVQVLTOLMPYGCIL 373
 QY 814 DHVENKRGRLGSDQLLWMCQIAKMSYLEDVRLVHRDLAARVLYKSPNRYKXITDFGLA 873
 Db 374 DYIREHKDNISQYLLNWCQVQIAKMSYLEDVRLVHRDLAARVLYKXITDFGLA 433
 QY 874 RLILIDETEVYADOGKXPIKMWLESLIRRTTHOSDWSYGVYTWELMTFGAKYDGIIP 933
 Db 434 KLLGADKEVYHABGKXPIKMWLESLIRRTTHOSDWSYGVYTWELMTFGAKYDGIIP 493
 QY 934 AREIPDLLEKEERLPDPPICTIDVYIMWKCMWIDSECRPRFRELVSERPMARPCQSFV 993
 Db 494 ASEISVYLEKERLPDPPICTIDVYIMWKCMWIDSECRPRFRELVSERPMARPCQSFV 553
 QY 994 VIO-Q-NEDLGPASPLDSTFTYRSLLEDMDGDLVAEEYLVPOGFCPPDPAFGAGMVRH 105
 Db 554 VIQDERMHLPSPDTSKFRYRLMEEDMEDIVDADEYLVPOGFF----- 598
 QY 1053 HRSSSTRSGGDLTLGLEPSEEEAPRSPL-----APSEAGSDVFDGDLGMAKGLQSL 110
 Db 599 -NSPT-----SRFLLSSLSATSNNSATNCID-----RNGQGH 631
 QY 1108 PTHDPSPLOQYSEDPYPLPSET--DGYAAPLTCSPQPEYVNPQDVAPPPSPREBPPLPA 116
 Db 632 PVREDSFVQRYSSDPGTFLEESIDGFL-----PAPEYVNPQ--LMPKKS----- 675
 QY 1166 ARPAGATLERAKTLSPGKGVKDVDF-----AFGAVENPEYL 1203
 Db 676 -----TANVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPXYL 715
 RESULT 11
 Q086714 PRELIMINARY; PRT; 567 AA.
 AC Q086714;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 CX NCB1_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennistrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";

OncoGene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; -.
 DR HSSP; P11362; 1FCG.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 KW NON_TER
 FT 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 2,3e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCVACAHYKDPFCVACSPGVKPDLSYPIWKFPDEGACQPCPINCTHSCVDL 643
 DB 1 GP--DHCMCAHFIIGPCHVCACFAGVLGENDTL-VMKYADANAVCQLCHNCTRGCKGP 57
 QY 644 DDGCPAEGRASPLSTISAVV-GILVVVVGVPGLIKRQCKIRKYMRLLOTEL 702
 DB 58 GLEGG--NSKRTPSIAGVVGGLCLVVGGLGILGILRR-HIVKRTLRLLQEREL 113
 QY 703 VEPLTPSGAMPNQAQMRILKETELRKVYLGSGAFGVYKGIWIPDENYKIPAIKVLK 762
 DB 114 VEPLTPSGAPNQAHLRIKETEFKVKVLSGAFGVYKGLWIPDEBKXIPAIKELR 173
 QY 763 EHTSPKAKEILDEYVWAGSPYVSLGICLTSTYQVLTQMPRGICLLDHVRENRGR 822
 DB 174 EHTSPKAKEILDEYVWASVDNPRVCLGICLTSTYQVLTQMPRGICLLDIYREKDN 233
 QY 823 LGSQDLNMCQIAKMGSYLEDVRLVHRDLAARVLYKSPNHVKITDFFGLARLIDETE 882
 DB 234 IGSQYLLNMCQIAKMGVYLEDERRLVHRDLAARVLYKTPHVKITDFFGLAKLGDKE 233
 QY 883 YHADGKVPYKMALESILRRRFTHQSDVSYGVTVWELMTFGAKPYDGIPAREIPDLE 942
 DB 294 YHAEQKVPYKMALESILHRIYTHQSDVSYGVTVWELMTFGAKPYDGIPAREIPDLE 353
 QY 943 KGERLPQPICTIDVYIMWKCMIDSECRPRELVSFSESRMARDQRVVO-NEDLG 1001
 DB 354 KGERLPQPICTIDVYIMWKCMIDADSRKRELIAEFSKARDPRLYLQGGERRH 413
 QY 1002 PASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSG 1061
 DB 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPOQGF-----NSPST--- 454
 QY 1062 GGDULTGLPESEERRSPL-----APSEGASVFPDGLGMAKAGLQSLPHDSPLQ 1116
 DB 455 -----SKTFLSSLSATSNNSATNCID-----RNGQGHPRVRESFVQ 491
 QY 1117 RYSEDPVLPDSET--DGYVAPLTCSPQPEYVNOVDVRRPSPHREGPLPAAPAGATLE 1174
 DB 492 RYSDPTGNFLSEIDQGL-----PAPEYVNO-LMPKKPS----- 526
 QY 1175 RAKTLSPKNGVVDVF-----AGGAVENPEYI 1203
 DB 527 ----TAMVQNOIYNNISITAIKSLPMDSRYONSHSTAVDNPYI 566

RESULT 12
 O64895 PRELIMINARY; PRT; 412 AA.
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.

GN PP1659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCB1_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Man D.F., Gu J.R.;
 RA "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318349; AAL55856.1; -.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR00719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 5.1e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 895 MALSELRRRTTHQSDVSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICT 954
 DB 1 MALESILRRRTTHQSDVSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICT 60
 QY 955 IDVYIMWKCMIDSECRPRELVSFSESRMARDQRVVO-NEDLG 1001
 DB 61 IDVYIMWKCMIDSECRPRELVSFSESRMARDQRVVO-NEDLG 1001
 QY 1015 LEDDDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTGLPSEE 107
 DB 121 LEDDDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTGLPSEE 180
 QY 1075 EAPSPPLAPSGASDVFPDGLGMAKAGLQSLPHDSPLQRYSEDPVLPDSETDGYV 113
 DB 181 EAPSPPLAPSGASDVFPDGLGMAKAGLQSLPHDSPLQRYSEDPVLPDSETDGYV 240
 QY 1135 APLTCSPQPEYVNOVDVRRPSPHREGPLPAAPAGATLERAKTLSPKNGVVDVFAPG 119
 DB 241 APLTCSPQPEYVNOVDVRRPSPHREGPLPAAPAGATLERAKTLSPKNGVVDVFAPG 300
 QY 1195 GAVENPEYIYTGGAAPQ-----HPPA---FSPAFLNL 122
 DB 301 GAVENPEYIYTGGAAPQ-----HPPA---FSPAFLNL 122
 QY 1227 YVMD-QDPPER-----GAPSTFGTPTAEN 1251
 DB 361 YVMD-QDPPER-----GAPSTFGTPTAEN 1251

RESULT 13
 O64895 PRELIMINARY; PRT; 962 AA.
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-v-erb-A-v-erb-B protein.
 GN GAG-V-ERB-A-V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
 OX NCB1_TaxId=11861;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=90206603; PubMed=1969616;
RA Brakin A., Jackson J., Bishop J.M., McCatley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B."
CC Oncogene 5:15-24(1990)
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA36459.1; -.
DR EMBL; X52211; CA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR000536; Hormone_rec_1ig.
DR InterPro; IPR001723; Scdhm_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Casteroid.
DR Pfam; PF00104; hormone_rec.1.
DR Pfam; PF00069; kinase_rec.1.
DR SMART; SM00105; ZF-C4.1.
DR PRINTS; PR00398; STRDHOMER.
DR PRINTS; PR00047; STROIDFINER.
DR ProDom; PD000001; Euk_kinase.1.
DR ProDom; PD000035; Znf_Casteroid.1.
DR SMART; SM00430; HOL1.1.
DR SMART; SM00219; TyKc.1.
DR SMART; SM00399; ZNF_C4.1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger
SQ
SEQUENCE 962 AA; 108320 MW; 3C5ED791EB95CE CRC64;

Query Match 24.1%; Score 1654; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 3.5e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 547 VEEGRVLOGLPRE YNNAR-HCLP-----CHPECO 574
DB 354 IEKQESYLAFFHYIYRKHNIPHEWSKILMKVADLRMGAYHASFIMKVECTEIS 413
QY 575 PONGSVTCFGEADQCAVCAHYKDPFCVACPSGVKPDLSYPIKFPDEEGACOPCI 634
DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACAGVLGENDTL-VKRYADANAVALCOLCHP 465
QY 635 NCHTSCVDLDKGCPCAPQASPLTISAVV-GILLVVLGVVFGILIKRQOKIRKXTM 693
DB 466 NCTRGCKGPGLEGP--NSGKTPSIAAGVGGILCLVAVGLGIGLYLRR-HIYKRTL 521
QY 694 RRLQETELVEPLTPSGAMPNOAMRILKETELRKVKVLGSGAFGTIVYKIMIPGSENVK 753
DB 522 RRLQEREIVLPLTPSEANNOAHLIKETEFKVKVLFQAFGVYVGLMIPSEKVT 581
QY 754 IPAVKVLRRENTSPKANKELIDEAYVMAVGSPYVSRLLIGICTSTVOLVQLMPYGLL 813
DB 582 IPAVIXELREATSPKANKELIDEAYVMAVDNPHVCRLLGICLTSTVOLITQLMPIYGLL 641
QY 814 DHVENRNGRLSGODLLNMCQIAKMSYLEDVVLRDLAARVLYKSNHVKITIDFGLA 873
DB 642 DYIREHKNIGSQILLNMCQIAKGMNYLEERHMYRDLAARVLYKTHQVKTIDFGLA 701
QY 874 RLIDIDETEHADGKVPKIKMALESILRRFTHSDVMSYGVTWELMTFGAKPYDGI 933
DB 702 KOLGADKEHYHAEGKVPKIKMALESILHRIYTHSDVMSYGVTWELMTFGSKYDGI 761
QY 934 AREPDLLEKGERLPQPICTIVYMIYKCMWIDSECPREPELVSESNARPORFV 993
DB 762 ASEISSVLEKGERLPQPICTIVYMIYKCMWGSADSPKRELLAEPSSKMARPPRYL 821
QY 994 VIQ-NEDLGPPASPLDSTFYRSLLEDMDGDLVAEEVLPVPOGFCFCDPAAGAGVHR 1052
DB 822 VIQDERMHLPSPTDSFRTLMEEEDMEDIVADAYLVPHQGF----- 866

```

```

QY 1053 HRSSSTSGGDLTLGLPSEEPREPLAPSGAGSDVFDGLGMAAKGLOSPTHDP 111
DB 867 -NSPST-----SRFLPSSLSATSN-----NSATYCIRNGGH-- 898
QY 1113 SPLORYSDEPTVPLPSTDTGVVAFLTCSPOPEYVNOGPDVPPSPREGEPLPARPAGAT 117
DB 899 -----PVREDGFL-----PAPEVYNQ--LMPKKPSTAMVQNIYNYISLT 936
QY 1173 -LERAKTLSPCKNGVNDVAFGAVENPEYL 1203
DB 937 AISHLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scolding P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation."
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA30024.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TyKc.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ
SEQUENCE 545 AA; 60899 MW; 140DCB8C0AF8AF4 CRC64;

Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 7.2e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GPEADQCAVCAHYKDPFCVACPSGVKPDLSYPIKFPDEEGACOPPINCHSCVDL 643
DB 1 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VKRYADANAVALCOLCHPCTRCCKGP 57
QY 644 DDKCPAEGRASPPLTISAVV-GILLVVLGVVFGILIKRQOKIRKXTMRLLQETEL 702
DB 58 GLBSCP--NSGKTPSIAAGVGGILCLVVGIGLYLRR-HIYKRTLRLLOERL 113
QY 703 VEPILTPSGAMPNOAMRILKETELRKVKVLGSGAFGTIVYKIMIPGSENVKIPVAIKVL 762
DB 114 VEPILTPSGEAPNOAMRILKETEFKVKVLFQAFGTIVYKIMIPGSEKVTIPVAIKEIR 173
QY 763 ENTPSPKANKELIDEAYVMAVGSPYVSRLLIGICTSTVOLVQLMPYGLLDHVENGR 822
DB 174 EATSPKANKELIDEAYVMAVDNPHVCRLLGICLTSTVOLITQLMPIYGLLDYIREHKN 233
QY 823 LSGODLLNMCQIAKMSYLEDVVLRDLAARVLYKSNHVKITIDFGLARLIDETE 882
DB 234 IGSQYLLNMCQIAKGMNYLEERHMYRDLAARVLYKTHQVKTIDFGLAQLCADEKE 293
QY 883 YHADGKVPKIKMALESILRRFTHSDVMSYGVTWELMTFGAKPYGIPAREPDLLE 942
DB 294 YHAEGKVPKIKMALESILHRIYTHSDVMSYGVTWELMTFGSKPYDGIASEISSYLE 353

```

```

QY 943 KGERLPQPICTIDYVMIMVKCMIDSCBPRPRELVESEFSRMARDQRFVYIO-NEDLG 1001
DB 354 KGERLPQPICTIDYVMIMVKCMIDSCBPRPRELVESEFSRMARDQRFVYIOGDRMR 413
QY 1002 PASPLDSTFYRSLLEDDMDLVDAEYLVPOQGFCEPCDPAPGAGVHHNRSSSTRSG 1061
DB 414 LPSPTDSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1062 GGDLTGLLEPSEEAAPRSL-----APSEGAGSDVFCGDLGMGAAGKQSLPTIDPSPLQ 1116
DB 455 -----SRTPULSSLSATSNNSKTNIDNKG-----H----- 461
QY 1117 RYSEDPVPLPSETDGYAFLTCSQPEYVQNPDPVROPSPSPREGPLPAPAPAGAT-LER 1175
DB 482 -----PYREDQFL-----PAPYVNO-LMPKPESTAVVQIYVYISLTASK 523
QY 1176 AKTLSPGKGVKVDVPAFGANENPEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Mahle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=LIVER;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateno Y., Nikaido T., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Morone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD4149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23668.1; -.
DR EMBL; AK004883; BAB23664.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; MGI:95294; Egr.
DR InterPro; IPR00494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep-Like; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 21.4%; Score 1464.5; DB 11; Length 655;
Best Local Similarity 43.4%; Pred. No. 4.4e-102;
Matches 281; Conservative 98; Mismatches 235; Indels 33; Gaps 10

```

```

QY 11 LILALPREGA--STOVCTGDMRLRLPASPTHLDLRLHYQCGOVVQGNLELTYPTN 68
DB 14 LITLCAAGGLAEKRYCGTGSNLTQLGTEDHFLSLQMYNCEVVLGNLELTYVQRN 73
QY 69 ASLSFLQDIQEVQGVYLIHNVQVRLRLRIYRGTQLFEDNYALAVLDNGPPLNNTTP 124
DB 74 YDLSFLKTIQVAGVLIALTVERILENLIQIRGNALYENTYALALISN----- 124
QY 129 VTGASPGRLREQLRLSLTELKGVVLQRPOLCYOQTILMKDI----FKNNOLALTLI 184
DB 125 -YGNRTGLRLPKMNOETILIGAVRSNPNPICNMPTIQMRDVLQNVFNSMSMDL--- 186
QY 185 DTRSRACHPCSPCKSGRCWSESEDCSLTRTVACGAGA-RCKGPLPTDCCGEGCAAG 243
DB 181 -QSHPSCKPKDPECPNSGSCWGGSEVNCQTLTKICAQGSHRCRGSPEPDCCHNCAAG 235
QY 244 CTGPHNSDCLACHFNHSIGTELHCPALVTYNTDFESMNPBGRYTFGASCYATCPYNY 303
DB 240 CTGPRESDCLVQCFODEATCKDCPPLMYPPTYQMDVNPBGRYSFGATCYVKKCPRNY 295
QY 304 LSTVGSCTLVCPILHNOEYTAEDGTQRCCKSPCARVYGLGW--FNNFTVFWLRLVPK 363
DB 300 VVTDHGSCTVRAAGDYIEV-EDDGRKCKCKDGPCKRYCNGIGIGEKD-TLS----- 350
QY 362 VSASHLEEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIRANSKFIIQITELGYLYISAW 421
DB 351 INATNIKFKKCYCTAISGLHILPAVAFGDSFTTRPPLDPRELEITLKVKEITGILLIQAW 410
QY 422 PDSLPDSVFNLOVIGRILHNVAVSLTIOGISLIGRSIRLSGLALHNNHTHC 481
DB 411 PDNWTDLAFENLILNGRTKQHQFSLAVAGLNTISLGRSLKEISDGVYIISGNRNLC 470
QY 482 FVHTVPMQDLFRNPHQALLHTANRPEDECEVGEGLAQHQLCARGHCWGPPTQCVNCSQFL 541
DB 471 YANTINWKKLFGTNGQTKIMNRAEKDCAVNVHVCNPLCSSGCMGPEPRDVCQGNNS 530
QY 542 RGQCEVEECRLQLPREYVNAARCLFCHPECQPNQSVTCGPEADQCVACHYKDPPE 601
DB 531 RGRCEVECNLTLEGEPEFENSECICHCPCFQANNITCTGRGPDNCIQCAHYIDGFH 590
QY 602 CVACPSGVKPDLSYMPIMKFPDEGACQCPINCTFSCTVDLDDKGC 648
DB 591 CVKICCPAGIMENNTL-VMKYADANNVCHLCHANCYIGCGAPLQGC 636

```

Search completed: July 22, 2003, 09:01:17
 Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGLDVP 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Genseq.101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	6661	97.2	1255	21	AA92620 Human heregulin 2
2	6661	97.2	1255	22	AAE12130 Human tyrosine kin
3	6661	97.2	1255	22	AA860167 Human transgene pla
4	6661	97.2	1255	23	AAU74545 Human HER2 (ErbB2)
5	6655	97.1	1255	17	AAU01111 HER-2/neu protein
6	6655	97.1	1255	20	AA92406 Human HER-2/neu on
7	6655	97.1	1255	21	AA821198 Human HER-2/neu pr
8	6655	97.1	1255	21	AA84780 Amino acid sequenc
9	6655	97.1	1255	22	AA85458 Human HER-2/neu pr
10	6655	97.1	1255	22	AA88267 HER2/neu amino aci

11	6655	97.1	1255	23	AAE24067 Human Her-2 protei
12	6655	97.1	1255	23	AAE20479 Human Her-2/neu pr
13	6655	97.1	1255	23	AA851143 Human Her-2/neu or
14	6655	97.1	1255	23	AAU77114 Human Her-2/neu pr
15	6616	96.5	1433	14	AA939568 Sequence of c-erbB
16	6491	94.7	1223	13	AAU98923 Human breast cance
17	6338	92.5	1200	21	AA821208 Human HER-2/neu pr
18	5863.5	85.6	1256	23	AA821199 Rat Her-2/neu prot
19	5863.5	85.6	1256	23	AA821199 Rat Her-2/neu onc
20	5842.5	85.3	1256	22	AA821206 Mouse Her-2/neu pr
21	5842.5	85.3	1256	22	AA821206 Amino acid sequenc
22	5842.5	85.3	1256	22	AA821206 Mouse Her-2/neu or
23	4741	69.2	919	23	AA821203 Human Her-2/neu fu
24	4741	69.2	919	23	AA821203 Her-2/neu extracel
25	3992.5	58.3	920	23	AA821203 Mouse Her-2/neu ex
26	3992.5	58.3	926	23	AA821204 Mouse Her-2/neu ex
27	3625	52.9	712	23	AA821204 Human HER-2/neu fu
28	3625	52.9	712	23	AA821204 Her-2/neu extracel
29	3479	50.8	782	18	AA821200 Her-2-GM-CSF immunc
30	3477	50.7	653	21	AA821200 Extracellular HER-
31	3477	50.7	653	23	AA821200 Human ErbB2 oncopr
32	3439	50.2	645	22	AA860408 Human ErbB2 extrac
33	3439	50.2	645	22	AA860408 DC8cFV-erbB2B2C fu
34	3374	49.2	951	21	AA821203 Human EGF receptor
35	3271	47.7	624	11	AA821203 Human Her-1 protei
36	3101	45.3	1210	21	AA819259 Human epidermal gr
37	3101	45.3	1210	21	AA819259 Amino acid sequenc
38	3101	45.3	1210	23	AA823019 Human EGF receptor
39	3101	45.3	1210	23	AA823019 Human Her-1 protei
40	3099	45.2	1210	22	AA868420 Human epidermal gr
41	3084	45.0	583	23	AA820483 Amino acid sequenc
42	3084	45.0	587	23	AA820481 Human protein for
43	3083	45.0	589	23	AA820482 Human protein for
44	3083	45.0	600	23	AA820482 Human protein for
45	3060	44.7	1210	23	AB851768 Human epidermal gr

ALIGNMENTS

RESULT 1	AA92620	standard; Protein; 1255 AA.
ID	AA92620	
AC	AA92620;	
XX	10-AUG-2000 (first entry)	
DT	Human heregulin 2 (Her2).	
XX	Human heregulin 2 (Her2).	
DE	Human heregulin 2 (Her2).	
XX	Human heregulin 2 (Her2).	
KW	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Homo sapiens.	
Key	Location/Qualifiers	
FT	1..173	
FT	/label= N-terminal	
FT	/note= "mature polypeptide"	
FT	5..25	
FT	/label= insertion_region	
FT	/note= "suitable for foreign epitope insertion"	
FT	59..73	
FT	/label= insertion_region	
FT	/note= "suitable for foreign epitope insertion"	
FT	103..117	
FT	/label= insertion_region	
FT	/note= "suitable for foreign epitope insertion"	
FT	149..163	
FT	/label= insertion_region	
FT	/note= "suitable for foreign epitope insertion"	
FT	174..323	
FT	Domain	

FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 359..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /label= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain
 PN MO200020027-A2.
 PD 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEB1-) M & E BIOTECH AS.
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX MPI: 2000-349917/30.
 DR N-PSDB; AAA09455.
 PT Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 XX
 XX Claim 62; Page 193-198; 220pp; English.
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Suddominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respectiv
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 CC
 SQ Sequence 1255 AA;
 Query Match 97.2%; Score 6661; DB 21; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1211; Conservative 10; Mismatches 13; Indels 8; Gaps
 QY 1 MELALCERWGLLALPPGASTQVCTGDKMLPASPETHLMHLVQGCQVVGSL 60
 DB 1 MELALCERWGLLALPPGASTQVCTGDKMLPASPETHLMHLVQGCQVVGSL 60
 QY 61 ELTYLPTNASSFLQDIQEVGYLIANNQVQVPLQRLIRIVRGSTQFEDNYALAVDNG 120
 DB 61 ELTYLPTNASSFLQDIQEVGYLIANNQVQVPLQRLIRIVRGSTQFEDNYALAVDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELRLSITELKKGVLIOGNPOLCYDTITLMKDFHKNQOLA 180
 DB 121 DPLNNTPTVTGASPGGLRELRLSITELKKGVLIOGNPOLCYDTITLMKDFHKNQOLA 180
 QY 181 LTLIDTNSRACHPSPCKSKSRCKGSSSEDCQSLRTVACGARGCKGPLPDDCCEQC 240
 DB 181 LTLIDTNSRACHPSPCKSKSRCKGSSSEDCQSLRTVACGARGCKGPLPDDCCEQC 240
 QY 241 AAGCTGPRHSDCLACLFHNSGICELHCALVTYNTDFFSMPNDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLACLFHNSGICELHCALVTYNTDFFSMPNDEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTVCPILHNOVTEDEDTQREKSKCAVCVGLGFNNFTYSFMLR-V 355
 DB 301 YNYLSTDVGSCTVCPILHNOVTEDEDTQREKSKCAVCVGLGFNNFTYSFMLR-V 355
 QY 360 PKVASHLERFAGCKKIFGSLAFLPESFDGPPASNTAPQYIKANSKFICTELTGYLYIS 419
 DB 360 PKVASHLERFAGCKKIFGSLAFLPESFDGPPASNTAPQYIKANSKFICTELTGYLYIS 419
 QY 420 AMPDGLPPLSVFQNLQVYRGRILNNGAYSLTLQGLGTSWGLRLSGGLALIHNNTH 479
 DB 420 AMPDGLPPLSVFQNLQVYRGRILNNGAYSLTLQGLGTSWGLRLSGGLALIHNNTH 479
 QY 480 LCFVHTVPMQDLFNPHQALLHTANRDEDECVGGLACHOLCARGCMWGGPQCVNCSG 539
 DB 480 LCFVHTVPMQDLFNPHQALLHTANRDEDECVGGLACHOLCARGCMWGGPQCVNCSG 539
 QY 540 FLRGOECVEECRVQLQI.PREVVNARHCLPCHPCCQPNQSVTCFGBADQVCAAHYKP 599
 DB 540 FLRGOECVEECRVQLQI.PREVVNARHCLPCHPCCQPNQSVTCFGBADQVCAAHYKP 599
 QY 594 PFCYARCPGSGKPLSLMPLMKPFDEBGAQCPINCHSCVLDKRGCPAERASPLTS 659
 DB 594 PFCYARCPGSGKPLSLMPLMKPFDEBGAQCPINCHSCVLDKRGCPAERASPLTS 659
 QY 660 IVSAVGLLVVVLGVVFGILIKRRQOKIKYTMRLLOETELVEPLTPSGAMPNQAKMR 719
 DB 660 IVSAVGLLVVVLGVVFGILIKRRQOKIKYTMRLLOETELVEPLTPSGAMPNQAKMR 719
 QY 720 ILKTELIRKXKV.GSGAGFVYVGIWIPQENKIVAIKVLRENTSPKANKILDEAVY 779
 DB 720 ILKTELIRKXKV.GSGAGFVYVGIWIPQENKIVAIKVLRENTSPKANKILDEAVY 779
 QY 714 ILKTELIRKXKV.GSGAGFVYVGIWIPQENKIVAIKVLRENTSPKANKILDEAVY 773

QY 780 MAGVSPYVSRLLGICLTSTVLTQVLTQMPYGGCLLDHVENRGLSGODLTMWCMQIAKGM 839
 DB 774 MAGVSPYVSRLLGICLTSTVLTQVLTQMPYGGCLLDHVENRGLSGODLTMWCMQIAKGM 833
 QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDPLARLLIDETEXHAODGKVPITMALES 899
 DB 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDPLARLLIDETEXHAODGKVPITMALES 893
 QY 900 ILRRRFTHQSVDWSTGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
 DB 894 ILRRRFTHQSVDWSTGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
 QY 960 IVKCMWIDSECRPPRFRELVSFESRMARDPQRFVYIQWEDIGPASPLDSTFYRSLLEDD 1019
 DB 954 IVKCMWIDSECRPPRFRELVSFESRMARDPQRFVYIQWEDIGPASPLDSTFYRSLLEDD 1013
 QY 1020 MGDLDVAEEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAPRS 1079
 DB 1014 MGDLDVAEEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAPRS 1073
 QY 1080 FLAPSEGAGSDVFDDDLGMGAAGKGIQSLPTHDPSPLOKYSDDPTVPLPSETDGYVADLTC 1139
 DB 1074 FLAPSEGAGSDVFDDDLGMGAAGKGIQSLPTHDPSPLOKYSDDPTVPLPSETDGYVADLTC 1133
 QY 1140 SPOPEYVNPQDVPRPQPSPREGRLPAAPAGATLERAKTILSPKXNGVYKVFARFAGAVEN 1199
 DB 1134 SPOPEYVNPQDVPRPQPSPREGRLPAAPAGATLERAKTILSPKXNGVYKVFARFAGAVEN 1193
 QY 1200 PEYLTPOGGAPOHPPPAPSPAFENLYWQDDPERGAPSTKGTPTAENPEYIGLDV 1259
 DB 1194 PEYLTPOGGAPOHPPPAPSPAFENLYWQDDPERGAPSTKGTPTAENPEYIGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 2
 AAE12130
 ID AAE12130 standard; Protein; 1255 AA.
 AC AAE12130;
 DT 18-DEC-2001 (first entry)
 DE Human tyrosine kinase-type receptor, HER-2.
 KM Therapeutic compound; major histocompatibility complex; vaccine;
 KM antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KM adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KM antigen presenting cell; human; tyrosine kinase-type receptor.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 PN WO200168677-A2.
 PD 20-SEP-2001.
 PF 16-MAR-2001; 2001WO-US40328.
 PR 16-MAR-2000; 2000US-0527487.
 PA (GENZ) GENZYME CORP.
 PI Nicolette CA;
 DR WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy; has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 PT
 PS Claim 4; Page 63-67; 69pp; English.
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 CC
 SQ Sequence 1255 AA;
 Query Match 97.2%; Score 6661; DB 22; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps ;
 QY 1 METALACGWLILALPPGASTQVCTGTDKRLRSPETHLDMLEHLVQCGVQVGNL 60
 DB 1 METALACGWLILALPPGASTQVCTGTDKRLRSPETHLDMLEHLVQCGVQVGNL 60
 QY 61 ELTYLPTNASLSPLODIOEVQYVLIANNQVQVPLQRLIRVGTQFLFNVALAVDNG 120
 DB 61 ELTYLPTNASLSPLODIOEVQYVLIANNQVQVPLQRLIRVGTQFLFNVALAVDNG 120
 QY 121 DPLNNTTVTGASFGGLRELQRLTEILKGGVLIQNPOLCYODITLMKDIFHXNOLA 180
 DB 121 DPLNNTTVTGASFGGLRELQRLTEILKGGVLIQNPOLCYODITLMKDIFHXNOLA 180
 QY 121 DPLNNTTVTGASFGGLRELQRLTEILKGGVLIQNPOLCYODITLMKDIFHXNOLA 180
 DB 121 DPLNNTTVTGASFGGLRELQRLTEILKGGVLIQNPOLCYODITLMKDIFHXNOLA 180
 QY 181 LTLIDNRSRACHPCSPKCKSRCKGSSSEDCQSLRTVCAAGGARGKGLPDDCCHEOC 240
 DB 181 LTLIDNRSRACHPCSPKCKSRCKGSSSEDCQSLRTVCAAGGARGKGLPDDCCHEOC 240
 QY 241 AAGCTGPHSDCLACLPHNSGICELCPALVTYNTDTPESMPREGRYTFGASCVTACP 300
 DB 241 AAGCTGPHSDCLACLPHNSGICELCPALVTYNTDTPESMPREGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTIWCPLHNOEVTAREDTQRCCKSKRCARVCYGLGMFNFTVSWLR-V 359
 DB 301 YNYLSTDVSGCTIWCPLHNOEVTAREDTQRCCKSKRCARVCYGLGMFNFTVSWLR-V 359
 QY 359 YNYLSTDVSGCTIWCPLHNOEVTAREDTQRCCKSKRCARVCYGLGMFNFTVSWLR-V 359
 DB 359 YNYLSTDVSGCTIWCPLHNOEVTAREDTQRCCKSKRCARVCYGLGMFNFTVSWLR-V 359
 QY 360 PKVASHLEEFAGCKKIFGSLAFPLESFDDCPASNTAPQYIKXANSKFIGTELTGLYIYS 419
 DB 354 RAYTSANIGFAGCKKIFGSLAFPLESFDDCPASNTAPQYIKXANSKFIGTELTGLYIYS 413
 QY 420 AMPDSDPLDSYFQQLQYIRGRILHNGAYSLTLOGLGISWGLRSLBELSGSGLAIHNNTH 479
 DB 414 AMPDSDPLDSYFQQLQYIRGRILHNGAYSLTLOGLGISWGLRSLBELSGSGLAIHNNTH 473
 QY 480 LCFVHTVPMQDLFNPFPQALHTANPREDECVBEGGLACHQLCARHGMGPGPTQCNCSQ 539
 DB 474 LCFVHTVPMQDLFNPFPQALHTANPREDECVBEGGLACHQLCARHGMGPGPTQCNCSQ 533
 QY 540 FLRGQECVBECKVYLGPREYVARNHCLPCHPECCOPONGSVTCFGEADQCVACAAYKXP 599
 DB 534 FLRGQECVBECKVYLGPREYVARNHCLPCHPECCOPONGSVTCFGEADQCVACAAYKXP 593
 QY 600 PFCVARGSGVKDPLSYMPIMKRPDEEGACQPPINCTHSQVLDLDDKGPAPQASPLTS 659
 DB 594 PFCVARGSGVKDPLSYMPIMKRPDEEGACQPPINCTHSQVLDLDDKGPAPQASPLTS 653

```
QY 660 IVSAVVGILLVVVLGVVFGILLIKRQOKIKRKYTWRLLOETELVEPLTPSGAMPNOAQR 719
DB 654 IVSAVVGILLVVVLGVVFGILLIKRQOKIKRKYTWRLLOETELVEPLTPSGAMPNOAQR 713
QY 720 ILKETELRKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779
DB 714 ILKETELRKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 773
QY 780 MAGVGSPIVSRLLIGLITSTVOLVTLQMPYGLDHPVENGRGLSGQLLWMCQIAKGM 839
DB 774 MAGVGSPIVSRLLIGLITSTVOLVTLQMPYGLDHPVENGRGLSGQLLWMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVUKSPNHVKITDPGLAFLDIDETEHADGKVPIMMALES 899
DB 834 SYLEDVRLVHRDLAARNVUKSPNHVKITDPGLAFLDIDETEHADGKVPIMMALES 893
QY 900 ILRRRFTHSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPDPPICITIVYM 959
DB 894 ILRRRFTHSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPDPPICITIVYM 953
QY 960 IMVKCMMIDSECRPRELIVSEFSFMSMADDPQRFVVIQNEDELGPASPLDSTFYRLLEDD 1019
DB 954 IMVKCMMIDSECRPRELIVSEFSFMSMADDPQRFVVIQNEDELGPASPLDSTFYRLLEDD 1013
QY 1020 MGDLYDAEEYLVPOQGFPCDPAPAGAGMVRHRSSTRSGGDLTLGLEPSEEARPRS 1079
DB 1014 MGDLYDAEEYLVPOQGFPCDPAPAGAGMVRHRSSTRSGGDLTLGLEPSEEARPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQLRYSDDPTVLPJSETDGVVAALTC 1139
DB 1074 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQLRYSDDPTVLPJSETDGVVAALTC 1133
QY 1140 SQPEYVNPDPVRPQPPSPREPLPAARPAGATLERATLSFGKXGVYKDYFAAGAVEN 1199
DB 1134 SQPEYVNPDPVRPQPPSPREPLPAARPAGATLERATLSFGKXGVYKDYFAAGAVEN 1193
QY 1200 PEYLTPQGGAAPQHPHPAFAFNLVYMDODPPERGAPSTFKGPTAENPEYLGLDV 1259
DB 1194 PEYLTPQGGAAPQHPHPAFAFNLVYMDODPPERGAPSTFKGPTAENPEYLGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S; Schwall R;
```

```
XX
DR MPI: 2001-061962/07.
DB N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
XX receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 97.2%; Score 6661; DB 22; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps :
```

1 MELALCRWGLLALPPGAASTVCTGTDMKRLPASPEHLDMLHLYOGCQVVGNL 60

1 MELALCRWGLLALPPGAASTVCTGTDMKRLPASPEHLDMLHLYOGCQVVGNL 60

61 ELTYVPTNASLFLDIOEVGYVLIANOVROVPLORLVRGTOLPEDNYALAVLDNG 126

61 ELTYVPTNASLFLDIOEVGYVLIANOVROVPLORLVRGTOLPEDNYALAVLDNG 126

121 DELNNTTPTVGTASPGRLRELQRLSTLEIKGVLIOGNPOLCYDITLMKDIFHKNNQLA 186

121 DELNNTTPTVGTASPGRLRELQRLSTLEIKGVLIOGNPOLCYDITLMKDIFHKNNQLA 186

181 LTLIDTNSRBAHPSPSPCKSRGMSSEDCSLRTVACAGGARCKGSLPTDCCHEOC 246

181 LTLIDTNSRBAHPSPSPCKSRGMSSEDCSLRTVACAGGARCKGSLPTDCCHEOC 246

241 AAGCTGPHGSDCLACTHNSGICELHCPALVTYNTDFTESMPNREGRYTFGASCVTACP 306

241 AAGCTGPHGSDCLACTHNSGICELHCPALVTYNTDFTESMPNREGRYTFGASCVTACP 306

301 YNYLSTDVGSCTLVCPHNOEVTAEDETQRECKSKPCARVCYGLGNFNTVSWLR-V 355

301 YNYLSTDVGSCTLVCPHNOEVTAEDETQRECKSKPCARVCYGLGNFNTVSWLR-V 355

360 PKVASHLSEFPAGCKITFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTGYLYIS 415

360 PKVASHLSEFPAGCKITFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTGYLYIS 415

420 AMPDSLPLSVFONLQVIRGRILNNGAYSLTLQGLISWLGSLRLSGGLALIHNNTH 479

420 AMPDSLPLSVFONLQVIRGRILNNGAYSLTLQGLISWLGSLRLSGGLALIHNNTH 479

444 AMPDSLPLSVFONLQVIRGRILNNGAYSLTLQGLISWLGSLRLSGGLALIHNNTH 473

480 LCFYHTVPMQDLFNPQHALLHTANRBEDEVGGLACHOLCARGHGWMGPGPOCVNCSG 539

480 LCFYHTVPMQDLFNPQHALLHTANRBEDEVGGLACHOLCARGHGWMGPGPOCVNCSG 539

474 LCFYHTVPMQDLFNPQHALLHTANRBEDEVGGLACHOLCARGHGWMGPGPOCVNCSG 533

540 FLRQGEVCEECRVLQGLPREVYNAHCLPCHPECOPOGNSVTCFGEPAOCVACAHYKXP 599

540 FLRQGEVCEECRVLQGLPREVYNAHCLPCHPECOPOGNSVTCFGEPAOCVACAHYKXP 599

534 FLRQGEVCEECRVLQGLPREVYNAHCLPCHPECOPOGNSVTCFGEPAOCVACAHYKXP 593

600 PFCVARGSGVCKPPLSYMPIMKPFDEGAQOCPINCTHSCVLDKXGCAEGRASPLTS 659

600 PFCVARGSGVCKPPLSYMPIMKPFDEGAQOCPINCTHSCVLDKXGCAEGRASPLTS 659

594 PFCVARGSGVCKPPLSYMPIMKPFDEGAQOCPINCTHSCVLDKXGCAEGRASPLTS 653

660 IVSAVVGILLVVVLGVVFGILLIKRQOKIKRKYTWRLLOETELVEPLTPSGAMPNOAQR 719

660 IVSAVVGILLVVVLGVVFGILLIKRQOKIKRKYTWRLLOETELVEPLTPSGAMPNOAQR 719

654 IVSAVVGILLVVVLGVVFGILLIKRQOKIKRKYTWRLLOETELVEPLTPSGAMPNOAQR 713

720 ILKETELRKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779

```

Db 714 ILKEELRKVKVLSGSAFGTVYKGIWPDGENVKIPVALILERENTSPKANELLDEAVV 773
Qy 780 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCLLDHYRENRGSLGSDLLNMOMQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCLLDHYRENRGSLGSDLLNMOMQIAKGM 833
Qy 840 SYLEDVLYHRDLAARVLYKSPNHVKITDPGLARLIDIDETEHADGGKVPKMALES 899
Db 834 SYLEDVLYHRDLAARVLYKSPNHVKITDPGLARLIDIDETEHADGGKVPKMALES 893
Qy 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRRLQPPICITIDVW 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRRLQPPICITIDVW 953
Qy 960 IMVKCMWIDSECRPRFELVSESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDD 1019
Db 954 IMVKCMWIDSECRPRFELVSESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDD 1013
Qy 1020 MGDLYDAEEYLVPOGGFCDDPAPGAGMWHHRSSSTRSGGDLTLGLESEEEAPRS 1079
Db 1014 MGDLYDAEEYLVPOGGFCDDPAPGAGMWHHRSSSTRSGGDLTLGLESEEEAPRS 1073
Qy 1080 PLASEGAGSDVFDGDLGMGAKGLQSLPTHDPQLQRYSEDETVPLPEETDGYAPLTC 1139
Db 1074 PLASEGAGSDVFDGDLGMGAKGLQSLPTHDPQLQRYSEDETVPLPEETDGYAPLTC 1133
Qy 1140 SPOPEYVNOQDVAPRPPSPREGEPLPAARPAATLEBAKTLSPGKGVKQVAPFGAVEN 1199
Db 1134 SPOPEYVNOQDVAPRPPSPREGEPLPAARPAATLEBAKTLSPGKGVKQVAPFGAVEN 1193
Qy 1200 PEYLTPOGGAPOPHPAPSPAFDNLXYWDOPPERGAPSPSTFKGTPTAENPEYLGDV 1259
Db 1194 PEYLTPOGGAPOPHPAPSPAFDNLXYWDOPPERGAPSPSTFKGTPTAENPEYLGDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-Apr-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
XX
PA 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
XX (SCHW/) SCHWALL R.
XX (SLIW/) SLIKOWSKI M.
XX

```

```

PI Erickson S, Schwall R, Slikowski M;
XX
DR WPI: 2002-163686/21.
XX
DR N-PDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
XX
Query Match 97.2%; Score 6661; DB 23; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps 2
Qy 1 MELAALCRWGLLALLPGASASTQVCTGDMKRLRPASPTHLDMLRHLYGQCQVVGNTL 60
Db 1 MELAALCRWGLLALLPGASASTQVCTGDMKRLRPASPTHLDMLRHLYGQCQVVGNTL 60
Qy 61 ELTYLPNASTLFDLQIOEVQGVLIYAHNQVQVPLQRIYRTGTLFEDNYALVLDNG 120
Db 61 ELTYLPNASTLFDLQIOEVQGVLIYAHNQVQVPLQRIYRTGTLFEDNYALVLDNG 120
Qy 121 DPLNNTPTVYASPGGLRELOLRSLTEILKGVLTQBNPOLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVYASPGGLRELOLRSLTEILKGVLTQBNPOLCYQDTILMKDIFHKNNOLA 180
Qy 121 DPLNNTPTVYASPGGLRELOLRSLTEILKGVLTQBNPOLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVYASPGGLRELOLRSLTEILKGVLTQBNPOLCYQDTILMKDIFHKNNOLA 180
Qy 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLRTVYAGGACARCKPPTDCHEQC 240
Db 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLRTVYAGGACARCKPPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACGFHNSGICELHCAVLTNTDPTESMPNEGRYTFASCVTACP 300
Db 241 AAGCTGPKHSDCLACGFHNSGICELHCAVLTNTDPTESMPNEGRYTFASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCELANQVETADGTQRCCKSKPCARVCYGLGFNNFTVSFWLR-V 359
Db 301 YNYLSTDVSGCTLVCELANQVETADGTQRCCKSKPCARVCYGLGFNNFTVSFWLR-V 359
Qy 301 YNYLSTDVSGCTLVCELANQVETADGTQRCCKSKPCARVCYGLGFNNFTVSFWLR-V 359
Db 301 YNYLSTDVSGCTLVCELANQVETADGTQRCCKSKPCARVCYGLGFNNFTVSFWLR-V 359
Qy 360 PKYSASHLEFPAGCKKIFESLAFLEBSEFGDPASTAPQYIKANSKFGITELTYLYIS 419
Db 360 PKYSASHLEFPAGCKKIFESLAFLEBSEFGDPASTAPQYIKANSKFGITELTYLYIS 419
Qy 354 RAVTSANIOEPAGCKKIFESLAFLEBSEFGDPASTAPQYIKANSKFGITELTYLYIS 413
Db 354 RAVTSANIOEPAGCKKIFESLAFLEBSEFGDPASTAPQYIKANSKFGITELTYLYIS 413
Qy 420 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGGISWLGRLSRLREGSGALIHNTHT 479
Db 420 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGGISWLGRLSRLREGSGALIHNTHT 479
Qy 414 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGGISWLGRLSRLREGSGALIHNTHT 473
Db 414 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGGISWLGRLSRLREGSGALIHNTHT 473
Qy 480 LCFVHTVPMDOULFRNPHQALHTANRPEDECVGEGSLAQHLCARGHCGPPTQCVNCSQ 539
Db 480 LCFVHTVPMDOULFRNPHQALHTANRPEDECVGEGSLAQHLCARGHCGPPTQCVNCSQ 539
Qy 540 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHYKDP 599
Db 540 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHYKDP 599
Qy 534 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHYKDP 593
Db 534 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHYKDP 593
Qy 600 PFCVACPSGVKPDLSYMPIMWKFDEBAGACPCPINTCTHSCVLDLDDKCAEQRASPLTS 659
Db 600 PFCVACPSGVKPDLSYMPIMWKFDEBAGACPCPINTCTHSCVLDLDDKCAEQRASPLTS 659

```

```

Db      594 PFCVACRCSGVPLDSTYMPIMKFPDEBACQPCINCINCHSVDDDDGCAEGRASLTLS 653
Qy      660 IVSAVAVGILLVVLGVVFGVGLIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNOQMR 719
Db      654 IVSAVAVGILLVVLGVVFGVGLIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNOQMR 713
Qy      720 ILKETELRKRVKLVSGAGFGTVKGIWIPDGENVAKIPIVAKVLRNTPSPXANKETLDEAVY 779
Db      714 ILKETELRKRVKLVSGAGFGTVKGIWIPDGENVAKIPIVAKVLRNTPSPXANKETLDEAVY 773
Qy      780 MAGVSPVYSRLGLCTLTSTVQLVTQMLPYCCLLDHYENRGRLGSDLLNMCQIAKGM 839
Db      774 MAGVSPVYSRLGLCTLTSTVQLVTQMLPYCCLLDHYENRGRLGSDLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGGKVPKMMALLES 939
Db      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGGKVPKMMALLES 893
Qy      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPICITIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPICITIDVYM 953
Qy      960 IMVKCMWIDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMWIDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLDVABEYLVPQGFPCPDPAAGAGMVRHRRSSSTRSGGDDTLGLPESEEARPS 1079
Db      1014 MGDLDVABEYLVPQGFPCPDPAAGAGMVRHRRSSSTRSGGDDTLGLPESEEARPS 1073
Qy      1080 PLABEGAGSDVDFGDDLGMAKAGLSLPHDPSLPLORYSEDPVPLPSTGDIYVAPLTC 1139
Db      1074 PLABEGAGSDVDFGDDLGMAKAGLSLPHDPSLPLORYSEDPVPLPSTGDIYVAPLTC 1133
Qy      1140 SPOEYVNVQPDVREPQPSPREGLPAARPAAGATLERAKTSLPGKNGVYKQVPAFGAVEN 1199
Db      1134 SPOEYVNVQPDVREPQPSPREGLPAARPAAGATLERAKTSLPGKNGVYKQVPAFGAVEN 1193
Qy      1200 PEYLTPOGGAAPQHPHPAPSPAFDNTLYWDODPEFGAPSPSTFKGTPTAENEYIGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPAPSPAFDNTLYWDODPEFGAPSPSTFKGTPTAENEYIGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 5
AA001111 standard; Protein; 1255 AA.
XX ID AA001111;
XX AC AA001111;
XX DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX KW breast cancer; ovary cancer; colon cancer; lung cancer;
XX KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key
XX FT 676..1255
XX FT Domain
XX FT /label= Intracellular domain
XX FT /note= "claimed domain, useful for immunisation"
XX FN W09630514-A1.
XX PD 03-OCT-1996.

```

```

PF 28-MAR-1996; 96WO-US01689.
XX 31-MAR-1995; 95US-0414417.
XX (UNIM) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX WPI: 1996-455361/45.
XX N-PSDB: AAT40739.
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX Claim 2; Page 56-61; 71pp; English.
XX Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.1%; Score 6655; DB 17; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
Qy 1 MELAALRWGILLALBPAGASTOVCTGDMKRLRASPETHIDMLRHLYQCGVQVGNL 60
Db 1 MELAALRWGILLALBPAGASTOVCTGDMKRLRASPETHIDMLRHLYQCGVQVGNL 60
Qy 61 ELTYLPTNASLSFLODIOEVQGVYLIAHQVQVPLQRLRIYRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLODIOEVQGVYLIAHQVQVPLQRLRIYRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTPVYASPGGRLRELQSLTEILKGGVLIQRNPQLCYQDTILMKDIFHXKNQLA 180
Db 121 DPLNNTPVYASPGGRLRELQSLTEILKGGVLIQRNPQLCYQDTILMKDIFHXKNQLA 180
Qy 121 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCSLTRTYCAGGACRCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCSLTRTYCAGGACRCKGPLPTDCCHQC 240
Qy 241 AAGCTGPHSDCLACIHFHNSGICEIHCPLALVTYNTDTESMPNPEGRTYFASCVTACP 300
Db 241 AAGCTGPHSDCLACIHFHNSGICEIHCPLALVTYNTDTESMPNPEGRTYFASCVTACP 300
Qy 241 AAGCTGPHSDCLACIHFHNSGICEIHCPLALVTYNTDTESMPNPEGRTYFASCVTACP 300
Db 241 AAGCTGPHSDCLACIHFHNSGICEIHCPLALVTYNTDTESMPNPEGRTYFASCVTACP 300
Qy 301 YNYLSTVGSCTLVCPILHNOEVAEDGTQRCCKSPCARVCYGLGMFNFTVSMFLR-V 359
Db 301 YNYLSTVGSCTLVCPILHNOEVAEDGTQRCCKSPCARVCYGLGMFNFTVSMFLR-V 359
Qy 301 YNYLSTVGSCTLVCPILHNOEVAEDGTQRCCKSPCARVCYGLGMFNFTVSMFLR-V 359
Db 301 YNYLSTVGSCTLVCPILHNOEVAEDGTQRCCKSPCARVCYGLGMFNFTVSMFLR-V 359
Qy 360 PKVASASLLEFFACCKKIFGSLAFPLSPFGDPAASNTAPQYIRANSKFTIGTELTYLVIS 419
Db 354 RAYTSANICIEFFACCKKIFGSLAFPLSPFGDPAASNTAPQYIRANSKFTIGTELTYLVIS 413
Qy 420 AMPDLSLPDLSVPQNLQVIRGRIIHNAYSLTIOGIGISWLGRLRELGGSLALHNHTH 479
Db 414 AMPDLSLPDLSVPQNLQVIRGRIIHNAYSLTIOGIGISWLGRLRELGGSLALHNHTH 473
Qy 480 LCFVHTVPMQDLFRNHQALLHTANRPEDECVGEGLAQCLCARGHCWGPPTQCVCNSQ 539
Db 474 LCFVHTVPMQDLFRNHQALLHTANRPEDECVGEGLAQCLCARGHCWGPPTQCVCNSQ 533
Qy 540 ELAAGECEECRYQLPREYVNAHRCCLPCHDECOFONGSVYCCFGPEADQCVACHYXDP 599
Db 534 ELAAGECEECRYQLPREYVNAHRCCLPCHDECOFONGSVYCCFGPEADQCVACHYXDP 593

```

```

QY 600 PCVAPCPGVKPDLSYPIKPFDEBAGCPCPICNTHSCVDLDKDGCAEORASPLTS 659
Db 594 PFCVAPCPGVKPDLSYPIKPFDEBAGCPCPICNTHSCVDLDKDGCAEORASPLTS 653
QY 660 IVSAVVGILLVVGIVVGLIKRQOKIRKYTMRLLOEELVEPLTPSGANPNQAQR 719
Db 654 IISAVGILLVVGIVVGLIKRQOKIRKYTMRLLOEELVEPLTPSGANPNQAQR 713
QY 720 ILKETELRKVKLGSGAGFVYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYV 779
Db 714 ILKETELRKVKLGSGAGFVYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYV 773
QY 780 MAGVSPYVSRLLGLCLTSTVQVLTQMPYGCILLDHRNNGRIGSODLLNMCQIAKGM 839
Db 774 MAGVSPYVSRLLGLCLTSTVQVLTQMPYGCILLDHRNNGRIGSODLLNMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALSS 899
Db 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALSS 893
QY 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIARIEIPDLLEKGERLPQPPICITIDVYM 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIARIEIPDLLEKGERLPQPPICITIDVYM 953
QY 960 IMVKCMIDSECRPRFRELVEFSRMARDPQRFVVIQNEDLGPASPLDSTFFYSLLDSD 1019
Db 954 IMVKCMIDSECRPRFRELVEFSRMARDPQRFVVIQNEDLGPASPLDSTFFYSLLDSD 1013
QY 1020 MGDLYDAEEYVLPQGFPCDPAPAGAGMVRHRRSSSTSGGDLTLGLEPSEBAPRS 1079
Db 1014 MGDLYDAEEYVLPQGFPCDPAPAGAGMVRHRRSSSTSGGDLTLGLEPSEBAPRS 1073
QY 1080 PLAPSEGASDVPFGDILGMAGAKGLSLPHDPEPLORVSEDDTVLPSTTDGYVAPLTC 1139
Db 1074 PLAPSEGASDVPFGDILGMAGAKGLSLPHDPEPLORVSEDDTVLPSTTDGYVAPLTC 1133
QY 1140 SPOEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKDVFAFGAVEN 1199
Db 1134 SPOEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKDVFAFGAVEN 1193
QY 1200 PEYLTPQGGAAPOHPPPAPSPADNLIYNDODPBERGAPSPFKGTPTAENDEYGLDV 1259
Db 1194 PEYLTPQGGAAPOHPPPAPSPADNLIYNDODPBERGAPSPFKGTPTAENDEYGLDV 1253
QY 1260 PV 1261
Db 1254 PV 1255

```

```

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
AC AAW92406;
XX
XX
XX 21-APR-1999 (first entry)
XX
XX Human HER-2/neu oncogene protein.
XX
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
XX malignancy; treatment; tumour.
XX
XX Homo sapiens.
XX
XX Key
XX Region 676..1255
XX /note= "region which elicits immune response"
XX
XX US869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIM ) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX MPI: 1999-152835/13.
XX N-PSDB; AAX01912.
DR
DR Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
XX
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
XX of this protein is used in a method for eliciting or enhancing an immune
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and
XX B cells to produce an immune response to the HER-2/neu protein. The
XX method can be used for immunisation against a malignancy in which the
XX HER-2/neu oncogene is associated and in the treatment of an existing
XX tumour, or to prevent tumour occurrence or reoccurrence.
XX
XX Sequence 1255 AA;
SQ
Query Match 97.1%; Score 6655; DB 20; Length 1255.
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2

```

```

QY 1 MELALCRWGLLALLPPGAASVQVCTGTDMLRLPASPETHLDMRLHYOGGVVQGNL 60
Db 1 MELALCRWGLLALLPPGAASVQVCTGTDMLRLASPETHLDMRLHYOGGVVQGNL 60
QY 61 ELYLPTNASTSIFQDIOEVQGVYLAHNOQVROPVQRLRIYRGTLFEDNYALAVLNG 120
Db 61 ELYLPTNASTSIFQDIOEVQGVYLAHNOQVROPVQRLRIYRGTLFEDNYALAVLNG 120
QY 121 DPLNNTPTVYGASPGGIREIQLSLTEILKGVYLIQRNPQLCYQDTILMKDIFHKNNQLA 180
Db 121 DPLNNTPTVYGASPGGIREIQLSLTEILKGVYLIQRNPQLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDNRSRAHPCHPSPCKGRKMGESSEDCSLRTRYCAGGCARCKGPLPTCCHEOC 240
Db 181 LTLIDNRSRAHPCHPSPCKGRKMGESSEDCSLRTRYCAGGCARCKGPLPTCCHEOC 240
QY 241 AAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDTFESMNPREGRTTFGASCTYACP 300
Db 241 AAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDTFESMNPREGRTTFGASCTYACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCGYGLMFNNFTVSEFLR-V 359
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCGYGLMEH-----LREV 353
QY 360 PKYSASHLEFPACCKKIFGSLATLPESFGDPASNTAPQYITANSKFICITLTGLYIS 419
Db 360 PKYSASHLEFPACCKKIFGSLATLPESFGDPASNTAPQYITANSKFICITLTGLYIS 413
QY 420 AMPDSLPLDSVFNLTQVIRGRILHNAAYSLTIQGLGISWLGRLSRLREGSGALHHNTH 479
Db 414 AMPDSLPLDSVFNLTQVIRGRILHNAAYSLTIQGLGISWLGRLSRLREGSGALHHNTH 473
QY 480 LCFVHTVPMQDLFRNPHQALLHTANPEDECVGEGIAHQQLCARGHCWPGPTQCVNSQ 539
Db 474 LCFVHTVPMQDLFRNPHQALLHTANPEDECVGEGIAHQQLCARGHCWPGPTQCVNSQ 533
QY 540 FLRGQECVEECRYLQGLPREYVNAHRCILCHPEQCPONSSVTCFPEPADQCYACAHYDP 599
Db 534 FLRGQECVEECRYLQGLPREYVNAHRCILCHPEQCPONSSVTCFPEPADQCYACAHYDP 593

```

```
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTS 659
D 594 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTS 653
QY 660 IVSAVVGILLVVLGVVFGIILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNOQMR 719
D 654 IISAVVGILLVVLGVVFGIILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNOQMR 713
QY 720 ILKETELRKVKVLGSAFGTYVKGIMIPDGENVKIPVAKYLRENTSKANKELIDEAYV 779
D 714 ILKETELRKVKVLGSAFGTYVKGIMIPDGENVKIPVAKYLRENTSKANKELIDEAYV 773
QY 780 MAGVSPYVSRLLGLICTSTVQLVTLQMPYGCILDHVENNGRLGSGODLIMWCMQIAKGM 839
D 774 MAGVSPYVSRLLGLICTSTVQLVTLQMPYGCILDHVENNGRLGSGODLIMWCMQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVLYKSPNNVKTIDFGLALILIDETETHAOGKVPIMWMALES 899
D 834 SYLEDVRLVHRDLAARNVLYKSPNNVKTIDFGLALILIDETETHAOGKVPIMWMALES 893
QY 900 ILRRRFTHSDVMSYGVTVMEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
D 894 ILRRRFTHSDVMSYGVTVMEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
QY 960 IMVKCMIMIDSECRPRERELVSFSFMAADPPQRFVITQMEDIGPASPLDSTYRSLLDDDD 1019
D 954 IMVKCMIMIDSECRPRERELVSFSFMAADPPQRFVITQMEDIGPASPLDSTYRSLLDDDD 1013
QY 1020 MGDVDAEEYLVPOOGFCPPDPAFGAGMVRHRSSTSGGGDLTLGLEPSEEEAPRS 1079
D 1014 MGDVDAEEYLVPOOGFCPPDPAFGAGMVRHRSSTSGGGDLTLGLEPSEEEAPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPDLQRYSEDPTVLPJSEDTGYVABLTC 1139
D 1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPDLQRYSEDPTVLPJSEDTGYVABLTC 1133
QY 1140 SPOPEYVNOPDVRPQPPSPREGPLPAARPAAGTILERAKTILSPGKNGYKDYFAFGAVEN 1199
D 1134 SPOPEYVNOPDVRPQPPSPREGPLPAARPAAGTILERAKTILSPGKNGYKDYFAFGAVEN 1193
QY 1200 PEYLPPOGGAAPQHPHPAFAFPAFNLVYWDQDPPERGAPSTFKGTPTAENPEYGLADV 1259
D 1194 PEYLPPOGGAAPQHPHPAFAFPAFNLVYWDQDPPERGAPSTFKGTPTAENPEYGLADV 1253
QY 1260 PV 1261
D 1254 PV 1255
```

RESULT 7
AAB21198
ID AAB21198 standard, protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu, oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN MO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000MO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX

```
PA (COR-) CORIXA CORP.  
PA (SMK) SMITHKLINE BEECHAM.  
XX  
PI Cheever MA, Gheyssen D;  
XX  
XX WPI; 2000-505976/45.  
DR N-PSDB; AAB69736.  
XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX  
PS Claim 52; Fig 7; 128pp; English.  
XX  
CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.  
CC  
SQ Sequence 1255 AA;  
Query Match 97.1%; Score 6655; DB 21; Length 1255;  
Best Local Similarity 97.4%; Pred. No. 0;  
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 1;  
QY 1 METALCMWGLLIALPPGAASTQVCTGTDKRLPASPTHMLMRLHYGCGVVGNTL 60  
D 1 METALCMWGLLIALPPGAASTQVCTGTDKRLPASPTHMLMRLHYGCGVVGNTL 60  
QY 61 ELTYLPTNALSFLQIDQEVQGYVLAHQVRYPLQRLRVGTQFEDNYALAVLDNG 126  
D 61 ELTYLPTNALSFLQIDQEVQGYVLAHQVRYPLQRLRVGTQFEDNYALAVLDNG 126  
QY 121 DPLNNTTPTVGTASPGAGBELQSLTELKGVLIQSNPQCYQDITLMQDIFKXNQLA 186  
D 121 DPLNNTTPTVGTASPGAGBELQSLTELKGVLIQSNPQCYQDITLMQDIFKXNQLA 186  
QY 181 LTLIDTNSRACHPCSPKCKSRCKGSESSDQSLRTVCAGCARGCKPLPTDCEHCQ 246  
D 181 LTLIDTNSRACHPCSPKCKSRCKGSESSDQSLRTVCAGCARGCKPLPTDCEHCQ 246  
QY 241 AAGCTGPKHSDCLACLAHNSGICELCPALVTYNTDTFESMPNBERYTFGASCCTACP 306  
D 241 AAGCTGPKHSDCLACLAHNSGICELCPALVTYNTDTFESMPNBERYTFGASCCTACP 306  
QY 301 YNYLSTDVGSCTLVCPILNQEVTAEDGTORCEKSKPCARVCYGLGEMNNFTVSFWLR-V 355  
D 301 YNYLSTDVGSCTLVCPILNQEVTAEDGTORCEKSKPCARVCYGLGEMNNFTVSFWLR-V 355  
QY 360 PKVSAASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGTELTGYLYIS 415  
D 360 PKVSAASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGTELTGYLYIS 415  
QY 420 AMPDCLPPLSVFQNLQVIRGRILNNGVYSITLQGLISWGLSLRSLRGLALIHNNTH 479  
D 420 AMPDCLPPLSVFQNLQVIRGRILNNGVYSITLQGLISWGLSLRSLRGLALIHNNTH 479  
QY 441 AMPDCLPPLSVFQNLQVIRGRILNNGVYSITLQGLISWGLSLRSLRGLALIHNNTH 473  
D 441 AMPDCLPPLSVFQNLQVIRGRILNNGVYSITLQGLISWGLSLRSLRGLALIHNNTH 473  
QY 480 LCFVHTVPMQDLFNNPQALHTANREDEDCVGBGLACHQLCARGHCGMGPPQVCNCSQ 539  
D 480 LCFVHTVPMQDLFNNPQALHTANREDEDCVGBGLACHQLCARGHCGMGPPQVCNCSQ 539  
QY 540 FLRGQECVEECRVYQGLPREYVNAHCLPCHPQCPONGSVTTFGBADQVCACATYKXP 599  
D 540 FLRGQECVEECRVYQGLPREYVNAHCLPCHPQCPONGSVTTFGBADQVCACATYKXP 599  
QY 534 FLRGQECVEECRVYQGLPREYVNAHCLPCHPQCPONGSVTTFGBADQVCACATYKXP 593  
D 534 FLRGQECVEECRVYQGLPREYVNAHCLPCHPQCPONGSVTTFGBADQVCACATYKXP 593  
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTS 659
```

```

Db      594 PFCVACPSGVKPDLSYMPIMKPFDEEGACQPCPINCTHSCVDLDKGCAPAGQASPLTS 653
QY      660 IVSAVVGILVVVYGVVVGILIKRRQOKIRKYIMRRLLOTELYEPLTPSGAMPNQAQR 719
Db      654 IISAVVGILVVVYGVVVGILIKRRQOKIRKYIMRRLLOTELYEPLTPSGAMPNQAQR 713
QY      720 ILKETELRKVKVLKSGAGFVYKGIWIPDGENVKIPAIIVYLRENTS PKAKKEILDEAVY 779
Db      714 ILKETELRKVKVLKSGAGFVYKGIWIPDGENVKIPAIIVYLRENTS PKAKKEILDEAVY 773
QY      780 MAGVSPYVSRLLIGICTSTVQVLTQMLPFGCLLDHRENRGRIGSODLLINMCQIAKGM 839
Db      774 MAGVSPYVSRLLIGICTSTVQVLTQMLPFGCLLDHRENRGRIGSODLLINMCQIAKGM 833
QY      840 SYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLIDETEVHADGGKVPKIMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLIDETEVHADGGKVPKIMMALES 893
QY      900 ILRRRFTHQSDVMSYGVTTWELMTFFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
Db      894 ILRRRFTHQSDVMSYGVTTWELMTFFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
QY      960 IMVKCMIMIDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYKSLLEDD 1019
Db      954 IMVKCMIMIDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYKSLLEDD 1013
QY      1020 MGDIVDAEYLVPOQGFPCDPAPAGAGMVAHHRSSSTRSGGDDLTLGLEPSEEBAPRS 1079
Db      1014 MGDIVDAEYLVPOQGFPCDPAPAGAGMVAHHRSSSTRSGGDDLTLGLEPSEEBAPRS 1073
QY      1080 PLASEGAGSVYFPGDILGMGAAGKLOSLPHHDSPFLORSYEDTVPPLPSTDDGYAPLTC 1139
Db      1074 PLASEGAGSVYFPGDILGMGAAGKLOSLPHHDSPFLORSYEDTVPPLPSTDDGYAPLTC 1133
QY      1140 SPQPEYVNPDPVRPQPSPREGSLPAARPAATLTERAKTILSPKNGVAVDVFAPFGAVEN 1199
Db      1134 SPQPEYVNPDPVRPQPSPREGSLPAARPAATLTERAKTILSPKNGVAVDVFAPFGAVEN 1193
QY      1200 PEYLTPOGGAAPQHPHPAPSPAFDNLYWDODPBERGAPPSFFKGTPTAENPEYIGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPAPSPAFDNLYWDODPBERGAPPSFFKGTPTAENPEYIGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AAy84780
ID      AAy84780 standard; Protein; 1255 AA.
XX
AC      AAy84780;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW      SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW      tumor cell proliferation; tissue degeneration; arthropathy;
KW      bone resorption; inflammatory disease; degenerative disorder;
KW      wound healing.
XX
OS      Homo sapiens.
XX
PN      W0200020579-A1.
XX
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99MO-CA00912.
XX
PR      02-OCT-1998; 98US-01.65192.
XX

```

```

PA      (UYWC-) UNIV MCMMASTER.
XX
PI      Muller WJ, Siegel PM;
XX
DR      WPI: 2000-303768/26.
XX
DR      N-PSDB; AAA14812.
XX
PT      Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX      erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erbB-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbB-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA;
XX
Query Match      97.1%; Score 6655; DB 21; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
QY      1 MELALGRWGILLALPPGAASTOVCTGDMKRLPASPEHLDMLRHLYOGQCVYQGNL 60
Db      1 MELALGRWGILLALPPGAASVOVCTGDMKRLPASPEHLDMLRHLYOGQCVYQGNL 60
QY      61 ELTYLPNNAISLFIQDIQEVQGYVLIANNOVROPVQRLRIYRGTLFEDNVALAVLNG 120
Db      61 ELTYLPNNAISLFIQDIQEVQGYVLIANNOVROPVQRLRIYRGTLFEDNVALAVLNG 120
QY      121 DPLNNTTPVYGASPGSLRELQSLLEILKAGVLIQRNPOLCYQDTILLMKDIFHKNNQLA 180
Db      121 DPLNNTTPVYGASPGSLRELQSLLEILKAGVLIQRNPOLCYQDTILLMKDIFHKNNQLA 180
QY      181 LTLIDINRSRACHPCSEMKGSRKWESESDQSLRTYVQAGZARCKGRLPTCCHEQC 240
Db      181 LTLIDINRSRACHPCSEMKGSRKWESESDQSLRTYVQAGZARCKGRLPTCCHEQC 240
QY      241 AAGCTGPHSDCLACLHFNHSGICEIHCPLVLYNTDFTESMNPBGRYTFGASCTYACP 300
Db      241 AAGCTGPHSDCLACLHFNHSGICEIHCPLVLYNTDFTESMNPBGRYTFGASCTYACP 300
QY      301 YNYLSTDVGSCTLVCPLHNOEVAEDGTQRCCKSKPCARVVCYGLGMFNNFTVSWLR-V 359
Db      301 YNYLSTDVGSCTLVCPLHNOEVAEDGTQRCCKSKPCARVVCYGLGMH-----LREV 353
QY      360 PKYSASLLEFPACCKIFGSLAFLPSPFGDPASNTAPQYIKANSKFIGITELVGLYIS 419
Db      354 RAYTSANIQEFACKKIFGSLAFLPSPFGDPASNTAPQPEQLQVFETLEIEITGLYIS 413
QY      420 AMPDSLDELVPONLQYIRGRILHNAYSITLQGLGISWLGARSUREIGSGALTIHHNTH 479
Db      414 AMPDSLDELVPONLQYIRGRILHNAYSITLQGLGISWLGARSUREIGSGALTIHHNTH 473
QY      480 LCPVHTVPMQDLFRNPHQALLHTANR PEDECVGEGLAGHQLCARGHCWPGPTQCVNCSQ 539
Db      474 LCPVHTVPMQDLFRNPHQALLHTANR PEDECVGEGLAGHQLCARGHCWPGPTQCVNCSQ 533
QY      540 FLRGQECVEECRYLQGLPREYNNARCLCHRECCQPNSSVTCFPEALDQCAACHYNDP 599
Db      534 FLRGQECVEECRYLQGLPREYNNARCLCHRECCQPNSSVTCFGEADQCAACHYNDP 593
QY      600 PFCVACPSGVKPDLSYMPIMKPFDEEGACQPCPINCTHSCVDLDKGCAPAGQASPLTS 659

```


DB 594 PFCVAPCPGKVPDLSTYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 653
QY 660 IVSAVVGILLVVLGVVFGILLIKRQOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 719
DB 654 IISAVVGLLVVLGVVFGILLIKRQOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 713
QY 720 IKKETELARKVVLGSGAGFTYKGIWIDGENVMI PVAIKYLENTPSPKANKELDEAYV 779
DB 714 IKKETELARKVVLGSGAGFTYKGIWIDGENVMI PVAIKYLENTPSPKANKELDEAYV 773
QY 780 MAGVSPYVSRLLGLITSTVQLVLTQMPYQCLLDHYRENGRLGSGDLLNMCQIAKGM 839
DB 774 MAGVSPYVSRLLGLITSTVQLVLTQMPYQCLLDHYRENGRLGSGDLLNMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGVPIKMALES 899
DB 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGVPIKMALES 893
QY 900 ILRRRFTHQSDVWSYGVTVWELMTFGANPYDGI PAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRRFTHQSDVWSYGVTVWELMTFGANPYDGI PAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVKCMIMIDSECRPFRELVESEFSRMAADPQRFVVIQNEDLGPRSLDSTFYSLLEDD 1019
DB 954 IMVKCMIMIDSECRPFRELVESEFSRMAADPQRFVVIQNEDLGPRSLDSTFYSLLEDD 1013
QY 1020 MGDLYDAEYLVPOQGFPCPDPAPOAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
DB 1014 MGDLYDAEYLVPOQGFPCPDPAPOAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1073
QY 1080 PLABSEGASDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPTVPLPSTDTGYVAPLTC 1139
DB 1074 PLABSEGASDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPTVPLPSTDTGYVAPLTC 1133
QY 1140 SPQPEYVNPQPVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKVPAFGAVEN 1199
DB 1134 SPQPEYVNPQPVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKVPAFGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPPPAFSPAFEDNLTYWDQDPPERGAAPPSTFKGPTAENPEYGLDV 1259
DB 1194 PEYLTPOGGAAPQHPPPAFSPAFEDNLTYWDQDPPERGAAPPSTFKGPTAENPEYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytosolic; vaccine; p185; c-erbB2.
OS Homo sapiens.
XX
PN WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX

PI Cheever MA, Hand-Zimmermann S;
XX WPI: 2001-476112/51.
DR N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 97.1%; Score 6655; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2

1 MELALCKWGLLLALLPPGAASVTCGTGMKRLPASPETH.DMLRHLYQGQVQVQGNL 60
1 MELALCKWGLLLALLPPGAASVTCGTGMKRLPASPETH.DMLRHLYQGQVQVQGNL 60
DB 1 MELALCKWGLLLALLPPGAASVTCGTGMKRLPASPETH.DMLRHLYQGQVQVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGVYLAAHNOVQVPLQRLRIVGTQLFEDNYALAVLNDG 120
DB 61 ELTYLPTNASLFLQDIOEVQGVYLAAHNOVQVPLQRLRIVGTQLFEDNYALAVLNDG 120
QY 121 DLNNTPPYVSGASGGRLQSLSTLTIKGVYLQGNPQLCYQDPTLLWQDIFHKNNQLA 180
DB 121 DLNNTPPYVSGASGGRLQSLSTLTIKGVYLQGNPQLCYQDPTLLWQDIFHKNNQLA 180
QY 121 DPLNNTTFFVVGASGGRLQSLSTLTIKGVYLQGNPQLCYQDPTLLWQDIFHKNNQLA 180
DB 121 DPLNNTTFFVVGASGGRLQSLSTLTIKGVYLQGNPQLCYQDPTLLWQDIFHKNNQLA 180
QY 181 LTLIDTRNSRACHPCSMCKSGRWCSSSDCCSLRTVQAGGACACCKGLPTDCHEQC 240
DB 181 LTLIDTRNSRACHPCSMCKSGRWCSSSDCCSLRTVQAGGACACCKGLPTDCHEQC 240
QY 241 AAGCTGPRGHSCLCLAFPHNSGICEHLCPALVYNYNDTFESMNPBGRYTFGASCTYACP 300
DB 241 AAGCTGPRGHSCLCLAFPHNSGICEHLCPALVYNYNDTFESMNPBGRYTFGASCTYACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEDETQRCCKSKPCARVCYGLMFPNFTVSFWLR-V 359
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEDETQRCCKSKPCARVCYGLMFPNFTVSFWLR-V 359
QY 360 PKVASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELGYLYIS 419
DB 360 PKVASHLEEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELGYLYIS 413
QY 420 AMPDLSPLSVFOQLVYIRGRIILHNGYSLTLGGLGYSWLGRLSELISGLALIHNNH 479
DB 420 AMPDLSPLSVFOQLVYIRGRIILHNGYSLTLGGLGYSWLGRLSELISGLALIHNNH 473
QY 474 LCPVHTVPMQDLFPNPHQALLHTANREDECEVGEGLACHQLCARGHGWCGRPTQCNCSG 533
DB 474 LCPVHTVPMQDLFPNPHQALLHTANREDECEVGEGLACHQLCARGHGWCGRPTQCNCSG 533
QY 540 FLRGQECVEECRYLQGLPREYVNAARHCLPHEPCQFQNGSVTFGEBAOQCVACAYKXP 599
DB 540 FLRGQECVEECRYLQGLPREYVNAARHCLPHEPCQFQNGSVTFGEBAOQCVACAYKXP 593
QY 600 PFCVAPCPGKVPDLSTYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 659
DB 594 PFCVAPCPGKVPDLSTYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 653
QY 660 IVSAVVGILLVVLGVVFGILLIKRQOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 719

```
DB 654 IISAVVGLLVVLGVFGILLIKRRQOKIRKXKTMRLIQETELVEPLTSPGAMENCAQMR 713
QY 720 IKETELRKVKVLGSGAGFTYVKGIIMPDGENVKIPVAIKYLRNTSPKANKETLDEAYV 779
DB 714 IKETELRKVKVLGSGAGFTYVKGIIMPDGENVKIPVAIKYLRNTSPKANKETLDEAYV 773
QY 780 MAGVSPVYSRLIGICLTSTVOLVTLQMPYGLLDHVENSGRLGSDLLNMCQIAKGM 839
DB 774 MAGVSPVYSRLIGICLTSTVOLVTLQMPYGLLDHVENSGRLGSDLLNMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDEGLAELIDETETHADGCKPIKMALES 899
DB 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDEGLAELIDETETHADGCKPIKMALES 893
QY 900 ILRRRFTHQSDVMSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVM 959
DB 894 ILRRRFTHQSDVMSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVM 953
QY 960 IMVKCMNIDSECRPRRELVSFESMARDPQRFVITQVEDLGASPLDSTYRSLLDDDD 1019
DB 954 IMVKCMNIDSECRPRRELVSFESMARDPQRFVITQVEDLGASPLDSTYRSLLDDDD 1013
QY 1020 MGDLYDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTIGLEPSEEAPRS 1079
DB 1014 MGDLYDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTIGLEPSEEAPRS 1073
QY 1080 PLAPSEAGASDVFDGDLGKAAGKIQSLPTHDPSFLQYSEDPTVPLPSETDGYVADLTG 1139
DB 1074 PLAPSEAGASDVFDGDLGKAAGKIQSLPTHDPSFLQYSEDPTVPLPSETDGYVADLTG 1133
QY 1140 SPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTSLPGKXGVYKDVAFAGAVEN 1199
DB 1134 SPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTSLPGKXGVYKDVAFAGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPERGAPSTFKGPTAENPEYGLDV 1259
DB 1194 PEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPERGAPSTFKGPTAENPEYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN MO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000MO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPI-M-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX
DR Keogh E;
XX
WPI; 2001-374995/39.
```

```
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 1999p; English.
PS
```

```
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
```

Sequence 1255 AA;

Query Match 97.1%; Score 6655; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps :

```
QY 1 MELALCRWGLLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYOGCGVQVGNL 60
DB 1 MELALCRWGLLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYOGCGVQVGNL 60
QY 61 ELTYLPTNASSFLQDIOEVGVYLAHNOVQVPLQRLIRVGTQIFEDNYALAVLDNG 124
DB 61 ELTYLPTNASSFLQDIOEVGVYLAHNOVQVPLQRLIRVGTQIFEDNYALAVLDNG 124
QY 121 DPLNNTTPTVGTASPGGRELQRLSTELILKGVYLIQNNPOLCYQDITLMDQIFKXNNOLA 180
DB 121 DPLNNTTPTVGTASPGGRELQRLSTELILKGVYLIQNNPOLCYQDITLMDQIFKXNNOLA 180
QY 181 LTLIDTNSRACHPCSPCKSGSRGSSSEDCQSLRTVACAGGACRCKGPLPTDCGHEQC 240
DB 181 LTLIDTNSRACHPCSPCKSGSRGSSSEDCQSLRTVACAGGACRCKGPLPTDCGHEQC 240
QY 241 AAGCTGPGHSCCLAHFNHSGICELCPALVYNTDFFSMPBERRYFFGASCTYACP 300
DB 241 AAGCTGPGHSCCLAHFNHSGICELCPALVYNTDFFSMPBERRYFFGASCTYACP 300
QY 301 YNYLSTDVGSCTLYCPLHNOEVTAEQTCRCKSKPCARVCYGLGWFNFTVSFWLR-V 359
DB 301 YNYLSTDVGSCTLYCPLHNOEVTAEQTCRCKSKPCARVCYGLGWFNFTVSFWLR-V 359
QY 360 PKVASHLEEFAGCKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGLYLYS 419
DB 360 PKVASHLEEFAGCKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGLYLYS 419
QY 354 RAVTSANIQEFAGCKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGLYLYS 413
DB 354 RAVTSANIQEFAGCKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGLYLYS 413
QY 420 AMPDLPDLASVFOQLQYIRGRIILHNGVYSLTLOGLISWGLSLRELSCGALLHHNTH 479
DB 420 AMPDLPDLASVFOQLQYIRGRIILHNGVYSLTLOGLISWGLSLRELSCGALLHHNTH 479
QY 414 AMPDLPDLASVFOQLQYIRGRIILHNGVYSLTLOGLISWGLSLRELSCGALLHHNTH 473
DB 414 AMPDLPDLASVFOQLQYIRGRIILHNGVYSLTLOGLISWGLSLRELSCGALLHHNTH 473
QY 480 LCFVHTVPMQDLFNNQOALLHTANRDECEVGEGLAQGLCARHGWGPGPTQVNCGQ 539
DB 480 LCFVHTVPMQDLFNNQOALLHTANRDECEVGEGLAQGLCARHGWGPGPTQVNCGQ 539
```

```

Db      474 LCFVHTVPMDOLEFRNPQALHTANRPEDECVSEGLACHQLCARGHCWGPPTQCVNCSQ 533
QY      540 FLRGQECVECECVLQGLPREYVNAHCLPCHPECCPQNGSVTTCGPEADQCVACAHTKDP 599
Db      534 FLRGQECVECECVLQGLPREYVNAHCLPCHPECCPQNGSVTTCGPEADQCVACAHTKDP 593
QY      600 PFCVAPCEGKVPDLSTYMPIMKFPDEGACQPCPINCCHSCVDLDDKCCPEQASPLTS 659
Db      594 PFCVAPCEGKVPDLSTYMPIMKFPDEGACQPCPINCCHSCVDLDDKCCPEQASPLTS 653
QY      660 IVSAVVGILLVVLGVNGLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQOMR 719
Db      654 IISAVVGILLVVLGVNGLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQOMR 713
QY      720 IKETELRKVKYLSGAFGTYYKGIWIDGENVKIPVAIKYLRNTSKANKELIDEAIV 779
Db      714 IKETELRKVKYLSGAFGTYYKGIWIDGENVKIPVAIKYLRNTSKANKELIDEAIV 773
QY      780 MAGVSPVYSRLGICLTSTVOLVTQMPYGLLDHVENRGLSGODLLNMCQIAKGM 839
Db      774 MAGVSPVYSRLGICLTSTVOLVTQMPYGLLDHVENRGLSGODLLNMCQIAKGM 833
QY      840 SYLEDVRLVHRLAARNVLKSPNHVKITDGLARLIDETEVHADGGKVPIMKMALES 899
Db      834 SYLEDVRLVHRLAARNVLKSPNHVKITDGLARLIDETEVHADGGKVPIMKMALES 893
QY      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
QY      960 IMVKCMIDSECRPFRELVSFSMAADPQRFVIONEDIGASPLDSTYRSLLLEDD 1019
Db      954 IMVKCMIDSECRPFRELVSFSMAADPQRFVIONEDIGASPLDSTYRSLLLEDD 1013
QY      1020 MGDLDVAEEYLVPOQGFPCPDPAFGAGMVAHRRSSSTRSGGDLTLGLEPSEEAPRS 1079
Db      1014 MGDLDVAEEYLVPOQGFPCPDPAFGAGMVAHRRSSSTRSGGDLTLGLEPSEEAPRS 1073
QY      1080 PLAPSEGAGSDVFDLDMGAAGKIGSIPTHDPSFLQYSEDPLVPJPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDLDMGAAGKIGSIPTHDPSFLQYSEDPLVPJPSETDGYVAPLTC 1133
QY      1140 SPQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTISFGKNGVVKCVAFAGAVEN 1199
Db      1134 SPQPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTISFGKNGVVKCVAFAGAVEN 1193
QY      1200 PEYLTPOGGAAPQHPHPAFSPAFPNLYYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPAFSPAFPNLYYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

```

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AAE24067;
XX
XX 23-SEP-2002 (first entry)
XX
XX Human Her-2 protein.
XX
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
XX
XX hyperproliferative disorder; prophylaxis; inflammation; antisense;
XX
XX tumour; gene therapy; phosphotriothate backbone.
XX
XX Homo sapiens.
XX
XX WO200222636-A1.
XX

```

PD      21-MAR-2002.
XX
XX PF 12-SEP-2001; 2001WO-US28572.
XX
XX PR 15-SEP-2000; 2000US-0663834.
XX
XX PA (ISIS-) ISIS PHARM INC.
XX
XX PI Bennett CF, Cowseert LM;
XX
XX WPI; 2002-471192/50.
XX
XX DR N-PSDB; AAD38904.
XX
XX PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor. Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans.
XX
XX PS Example 13; Page 95-107; 116pp; English.
XX
XX CC The invention relates to antisense compounds targeted to a nucleic
XX acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX CC that specifically hybridises with and inhibits the expression of Her2.
XX CC Antisense compounds of the invention are used for treating diseases or
XX CC conditions associated with Her2 such as hyperproliferative disorders
XX CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX CC neural or cardiac cancer. They are also useful prophylactically e.g.
XX CC to prevent or delay infection, inflammation and tumour formation. The
XX CC invention is also used in gene therapy. The present sequence is human
XX CC Her-2 protein.
XX
XX SQ Sequence 1255 AA;
XX
XX Query Match 97.1%; Score 6655; DB 23; Length 1255;
XX Best Local Similarity 97.4%; Pred. No. 0;
XX Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2

```

Db	534	FLRGCEVCESCRVLQGLPREYVNAHCLPCHPEQOPONGSTCGPGEADCCVCAHYKPF	593
Qy	600	PCVVARCPGKVPDLSTYMPIMKFPDEEGACOPPCINCHSCVDLDDGCPABOASPLTS	659
Db	594	PCVVARCPGKVPDLSTYMPIMKFPDEEGACOPPCINCHSCVDLDDGCPABOASPLTS	653
Qy	660	IVSAVVGILLVAVGVGFIILIKRQOKIRKYNRLLOETELVEPLTPSGAMPNOAKR	719
Db	654	IISAVVGILLVAVGVGFIILIKRQOKIRKYNRLLOETELVEPLTPSGAMPNOAKR	713
Qy	720	ILKETELRKVKVLSGAGFTYKGIWIPDGENVYIPAIIVLEBNTSPKXKEIIDEAVY	779
Db	714	ILKETELRKVKVLSGAGFTYKGIWIPDGENVYIPAIIVLEBNTSPKXKEIIDEAVY	773
Qy	780	MAGVSPYVSRLLGICLTSTVQVLTOLMPYGCCLLDHVENRGRVLSQDLINMCQIAKGM	839
Db	774	MAGVSPYVSRLLGICLTSTVQVLTOLMPYGCCLLDHVENRGRVLSQDLINMCQIAKGM	833
Qy	840	SYLEDVRLVHRDLAARNVLSBNHVKITDFGLARLLDIDETEHADGGKVPKMMALIS	899
Db	834	SYLEDVRLVHRDLAARNVLSBNHVKITDFGLARLLDIDETEHADGGKVPKMMALIS	893
Qy	900	ILRRRFTHQSDVMSYGYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPCTIDVYM	959
Db	894	ILRRRFTHQSDVMSYGYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPCTIDVYM	953
Qy	960	IMVKCMWIDSECRREFELVSEFSRMAKDPQRFVIONEDLGASPLDSTFYRSLLEDOD	1019
Db	954	IMVKCMWIDSECRREFELVSEFSRMAKDPQRFVIONEDLGASPLDSTFYRSLLEDOD	1013
Qy	1020	MGDLVDAEEYLVPOGFFCPBPAPAGAGM/HHRSSSTSSGGDLTLGLEPSEEPAPS	1079
Db	1014	MGDLVDAEEYLVPOGFFCPBPAPAGAGM/HHRSSSTSSGGDLTLGLEPSEEPAPS	1073
Qy	1080	PLAPSEAGSDVPDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTC	1139
Db	1074	PLAPSEAGSDVPDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTC	1133
Qy	1140	SPOPEYVNOPDVRRQPSRPGPLPARBAGATLEBPKTILSPKXGVVADVAFGGAIVEN	1199
Db	1134	SPOPEYVNOPDVRRQPSRPGPLPARBAGATLEBPKTILSPKXGVVADVAFGGAIVEN	1193
Qy	1200	PEYLLTPQGAAPQHPFPAPSPAFDNLVWDQPPPRGAPPSFTFKGTPTAEKPEYLLGDV	1259
Db	1194	PEYLLTPQGAAPQHPFPAPSPAFDNLVWDQPPPRGAPPSFTFKGTPTAEKPEYLLGDV	1253
Qy	1260	PV 1261	
Db	1254	PV 1255	

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.

XX AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human Her-2/neu protein; immune response; gene therapy; breast cancer;
XX human leukocyte antigen; HLA; vaccine; maligancy; cycostatic.
XX Homo sapiens.
XX Key
XX Region
XX Location/Qualifiers
XX 1021..1030
XX /note= "Naturally processed HLA-B*44-restricted epitope"
XX W0200214503-A2.

PD	21-FEB-2002.	
XX		
PF	14-AUG-2001; 2001WO-US41733.	
XX		
PR	14-AUG-2000; 2000US-225152P.	
PR	28-SEP-2000; 2000US-236428P.	
PR	21-FEB-2001; 2001US-270520P.	
XX		
PA	(CORI-) CORIXA CORP.	
PI	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;	
PI	McNeill PD, Vedyck TS;	
DR	WPI; 2002-280758/32.	
DR	N-PSDB; AAD32743.	
PT	Novel isolated Her-2/Neu polypeptide composition useful for therapy,	
PT	prevention and diagnosis of cancer, preferably breast cancer	
XX		
PS	Disclosure; Page 114-117; 123pp; English.	
XX		
CC	The invention relates to an isolated Her-2/Neu polypeptide composition effective for eliciting an immune response. The invention is useful for eliciting an immune response in a patient, where the patient is human leukocyte antigen (HLA)-B*44 positive or is affected with breast cancer. The composition is useful for the therapy and diagnosis of cancer, CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine and other compositions, for the diagnosis, prevention and treatment of human malignancies, for stimulating and/or expanding T cells specific for Her-2/Neu polypeptide and for inhibiting the development of cancer in a patient. The invention is useful for stimulating a T cell response in a human patient, as probe or primer for nucleic acid hybridization, to CC selectively form duplex molecules with complementary stretches of the entire Her-2/Neu gene or gene fragments of interest, to isolate a full CC length gene from a suitable library, and to direct expression of a CC polypeptide in appropriate host cells. The composition is useful in CC prophylactic or therapeutic applications and for the treatment of cancer, CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-associated malignancies. The invention is useful in gene therapy. The CC present sequence is human Her-2/neu protein.	
SO	Sequence	1255 AA;
Qy	Query Match	97.1%; Score 6655; DB 23; Length 1255;
Db	Best Local Similarity	97.4%; Pred. No. 0;
Db	Matches 1229; Conservative 11; Mismatch 14; Indels 8; Gaps 2	
Qy	1	MELALCRMGILLALLPAGASTQVCTGDMKRLRPASPEHLDMRHLYOGCQVVGNTL 60
Db	1	MELALCRMGILLALLPAGASTQVCTGDMKRLRPASPEHLDMRHLYOGCQVVGNTL 60
Qy	61	ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
Db	61	ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
Qy	121	DLNNTTPYTGASPGGLRELQRLSTLEIKSGVLIQRNPOLCYQDTIIMKDFHKNOLA 180
Db	121	DLNNTTPYTGASPGGLRELQRLSTLEIKSGVLIQRNPOLCYQDTIIMKDFHKNOLA 180
Qy	181	LTLIDNRSRACHPCSPMCKGSRCMGSESEDCQSLTRTYCAGGACRGPLPTDCHEQC 240
Db	181	LTLIDNRSRACHPCSPMCKGSRCMGSESEDCQSLTRTYCAGGACRGPLPTDCHEQC 240
Qy	241	AAAGCTPKASDCLACHFNHSGICELHCPALVTYNTDPESMNPDEGRYTFGASCVTACP 300
Db	241	AAAGCTPKASDCLACHFNHSGICELHCPALVTYNTDPESMNPDEGRYTFGASCVTACP 300
Qy	301	YNYLSTDVGSCTLVCEPLHNQVETAEADGTORCEKSKPCARVCYGVGMNMTVFWMR-V 359
Db	301	YNYLSTDVGSCTLVCEPLHNQVETAEADGTORCEKSKPCARVCYGVGMNMTVFWMR-V 359
Qy	360	PKYSASHLEEPGCKKIFGSLAFLEDSFDGDPASNTAPQYTKANSKFIGITELTGYLYIS 419
Db	360	PKYSASHLEEPGCKKIFGSLAFLEDSFDGDPASNTAPQYTKANSKFIGITELTGYLYIS 419

```

Db 354 RAVTSAHQEFAAGCKKIFGSLAFLPESEFDGDPASNTAPLQPEQLQVETLEITGYLYIS 413
Qy 420 AMPDSLPLSVFQNLQVIRGRILHNHGAASLTLOGLIGISWLGRLSRLSGLALIHNNTH 479
Db 414 AMPDSLPLSVFQNLQVIRGRILHNHGAASLTLOGLIGISWLGRLSRLSGLALIHNNTH 473
Qy 460 LCFVHTVMDQLFNNHQAALLHTANRPDECEVGGGLACHQICARGHCWGPPTQVNCQ 539
Db 474 LCFVHTVMDQLFNNHQAALLHTANRPDECEVGGGLACHQICARGHCWGPPTQVNCQ 533
Qy 540 FLRQGECEVECRVLOGLPREYVYNAHCLPCHPEQOPONGSVCGPPEADQVCAHYK 599
Db 534 FLRQGECEVECRVLOGLPREYVYNAHCLPCHPEQOPONGSVCGPPEADQVCAHYK 593
Qy 600 PFCVACPSGVKPDLSYMPWKPFDEEGACQPCINCTHSCVDLDKDGCAEGRASPLTS 659
Db 594 PFCVACPSGVKPDLSYMPWKPFDEEGACQPCINCTHSCVDLDKDGCAEGRASPLTS 653
Qy 660 IVSAVVGILLVVLGVVFGILLIKERQOKIRKYIMRLLQETELVEPLTPSGAMPNQAQR 719
Db 654 IVSAVVGILLVVLGVVFGILLIKERQOKIRKYIMRLLQETELVEPLTPSGAMPNQAQR 713
Qy 720 ILKETELRKVYVLSGAGFVYKGIWIPDGENVKIPVAIKVLEBNTSPKANKSILDEAY 779
Db 714 ILKETELRKVYVLSGAGFVYKGIWIPDGENVKIPVAIKVLEBNTSPKANKSILDEAY 773
Qy 780 MAGVSPVYSRLGLCTSTYQVLTQMLPYGCLLDHYRENRGRLSGODLIMWCQIAK 839
Db 774 MAGVSPVYSRLGLCTSTYQVLTQMLPYGCLLDHYRENRGRLSGODLIMWCQIAK 833
Qy 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMALES 899
Db 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMALES 893
Qy 900 ILRRFTHQSDVMSYGTWELMTFGAKPYDGIIPAREIPDLEKGRRLPQPICTIDVYM 959
Db 894 ILRRFTHQSDVMSYGTWELMTFGAKPYDGIIPAREIPDLEKGRRLPQPICTIDVYM 953
Qy 960 IMVCKMIDSECRFPREIVSEFRMARDPQFVYVONELGASPLDSTFYRSLLIEDD 1019
Db 954 IMVCKMIDSECRFPREIVSEFRMARDPQFVYVONELGASPLDSTFYRSLLIEDD 1013
Qy 1020 MGDIVDAEYVLPQGGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPARS 1079
Db 1014 MGDIVDAEYVLPQGGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPARS 1073
Qy 1080 PLASEGAGSVFPGDLMGAAGKLOSLPTHDSPLQRYSEDTVPLPSETDGYVAPLTC 1139
Db 1074 PLASEGAGSVFPGDLMGAAGKLOSLPTHDSPLQRYSEDTVPLPSETDGYVAPLTC 1133
Qy 1140 SPOPEYVNPQDVRPQPSPREGFLPARPAGATILERAKTILSPKNGVMDVFAFGAVEN 1199
Db 1134 SPOPEYVNPQDVRPQPSPREGFLPARPAGATILERAKTILSPKNGVMDVFAFGAVEN 1193
Qy 1200 PEYITPQGAAPQHPAPSPAFNDNYWDQPPERGAAPSTFFKCTPAENPEYIGLDV 1259
Db 1194 PEYITPQGAAPQHPAPSPAFNDNYWDQPPERGAAPSTFFKCTPAENPEYIGLDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

```

```

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
XX
XX Key
XX FH Location/Qualifiers
XX Domain 1..653
XX /note= "extracellular domain"
XX FT 676..1255
XX FT /note= "intracellular domain"
XX FT 900..1255
XX FT /note= "phosphorylation domain"
XX
XX MO200212341-A2.
XX
XX 14-FEB-2002.
XX
XX 03-AUG-2001; 2001WO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX
XX (CORI-) CORIXA CORP.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Cheever MA, Gheyssen D;
XX
XX WP1; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
XX or enhancing an immune response to the protein, has Her-2/neu
XX extracellular domain fused to Her-2/neu intracellular or
XX phosphorylation domain
XX
XX Claim 68; Fig 7; 14tp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to EGFR. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines
XX comprising the fusion proteins or nucleic acid molecules. In
XX preferred fusion proteins, the extracellular domain of a Her-2/neu
XX protein is fused to a Her-2/neu intracellular domain or
XX phosphorylation domain (or its DeltaPp fragment). An immune
XX response to Her-2/neu protein is elicited or enhanced by
XX administering the fusion protein in the form of a vaccine, or by
XX transfecting cells of an animal ex vivo with a nucleic acid
XX encoding the fusion protein, and delivering the transfected cells
XX to the animal. The fusion proteins, nucleic acids, and isolated
XX specific T-cells are useful for inhibiting the development of a
XX cancer, especially breast, ovarian, colon, lung or prostate cancer
XX in a patient. T cells that specifically react with a Her-2/neu
XX fusion protein can be used to remove tumour cells from a sample in
XX order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA.
XX
XX Query Match 97.1%; Score 6655; DB 23; Length 1255;
XX Best Local Similarity 97.4%; Pred. No. 0;
XX Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2.

```

```

Qy 1 METALACRMGLTALLPAGASNOVGTGDMRLRLPASEFTLDMRLYQCCVVOGNTL 60
Db 1 METALACRMGLTALLPAGASTOCTGDMRLRLPASEFTLDMRLYQCCVVOGNTL 60

```

```

QY 61 ELTYLPFTNASLSFIQDIOEVOGYVLIHNNQVRYPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPFTNASLSFIQDIOEVOGYVLIHNNQVRYPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASGGI RELQRLSTELIKGVYLIQNNPOLQVODITLMMDITHKNNOLA 180
DB 121 DPLNNTPTVTGASGGI RELQRLSTELIKGVYLIQNNPOLQVODITLMMDITHKNNOLA 180
QY 181 LTLIDITNRSRACHFCSPMCKGSRGSGSSSEDCQSLTRTVACAGGACAKGFLPTDCHEQC 240
DB 181 LTLIDITNRSRACHFCSPMCKGSRGSGSSSEDCQSLTRTVACAGGACAKGFLPTDCHEQC 240
QY 241 AAGCTGPRGSDCLCLHFNHSGICELHCPALVTYNTDFFESMNPESGRYTFGASCTYACP 300
DB 241 AAGCTGPRGSDCLCLHFNHSGICELHCPALVTYNTDFFESMNPESGRYTFGASCTYACP 300
QY 301 YNYLSTDVSGCTLVCPHNOEVTAEEDGTORCEKSKPCAVCYGLMFNNFTVSFWLR-V 359
DB 301 YNYLSTDVSGCTLVCPHNOEVTAEEDGTORCEKSKPCAVCYGLMFNNFTVSFWLR-V 359
QY 360 PKVASHLEEFAGCKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFIQTTELTYLYIS 419
DB 360 PKVASHLEEFAGCKKIFGSLAFLPESFGDDPASNTAPQYIKANSKFIQTTELTYLYIS 419
QY 420 AMPDLPDLVSFQNLQYIRGRILHNGAYSLTLQQLG:SWLGLRSLRELSGLALIHNNH 479
DB 420 AMPDLPDLVSFQNLQYIRGRILHNGAYSLTLQQLG:SWLGLRSLRELSGLALIHNNH 479
QY 480 LCFVHTVPMQDLFNNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPPTQVCNCSQ 539
DB 480 LCFVHTVPMQDLFNNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPPTQVCNCSQ 539
QY 540 FLRGQECVEBCRVLOGLPREYVNAHCLPCHPECOPONGSVTFEGEAAOCVACAHYKP 599
DB 540 FLRGQECVEBCRVLOGLPREYVNAHCLPCHPECOPONGSVTFEGEAAOCVACAHYKP 599
QY 594 PFCVAPCPSPGVKPDLSYMPIMKFPDEEGACQPCPINCHTSCVDLDKGPACRASPILTS 653
DB 594 PFCVAPCPSPGVKPDLSYMPIMKFPDEEGACQPCPINCHTSCVDLDKGPACRASPILTS 653
QY 660 IVSAVVGILLVYLVGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNOQMR 719
DB 660 IVSAVVGILLVYLVGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNOQMR 719
QY 720 ILKETELRKVYLGSGAFGVYKGIWIPDGENVKITVAIKVLRNENSPANKKEIIDEAV 779
DB 720 ILKETELRKVYLGSGAFGVYKGIWIPDGENVKITVAIKVLRNENSPANKKEIIDEAV 779
QY 774 MAGVGSFYVSRLLGICITSTVQLVTOQLMPYGCLLDHVRENRGRLGQDLLNMCQIAGM 833
DB 774 MAGVGSFYVSRLLGICITSTVQLVTOQLMPYGCLLDHVRENRGRLGQDLLNMCQIAGM 833
QY 840 SYLEDVRLVHRDLAARVLYKSPNHVKITDPGLARLLDIDEHYHADGGKVIKMMALES 899
DB 840 SYLEDVRLVHRDLAARVLYKSPNHVKITDPGLARLLDIDEHYHADGGKVIKMMALES 899
QY 894 ILRRRFTHQSDVMSYGVYTWELMTFGAKPYDGI:PAEIPDLLEKERRLPQPPICITIDVM 959
DB 894 ILRRRFTHQSDVMSYGVYTWELMTFGAKPYDGI:PAEIPDLLEKERRLPQPPICITIDVM 959
QY 960 IMYKCMWIDSECRPRELVESESRMAKRPQRRVVIQNDLGPASPLDSTFRLSLEDD 1019
DB 960 IMYKCMWIDSECRPRELVESESRMAKRPQRRVVIQNDLGPASPLDSTFRLSLEDD 1019
QY 1020 MGDLDVAEEYLVPOQGFCEPAPAGAGWVHHRRSSSTRSGGDLTLGLESESEAPRS 1079
DB 1020 MGDLDVAEEYLVPOQGFCEPAPAGAGWVHHRRSSSTRSGGDLTLGLESESEAPRS 1079
QY 1080 PLAPSEGAGSDVDGDLGMAAGLQSLTTHDPSLQRYSEDPYVPLDSETDGYVAPLTC 1139
DB 1080 PLAPSEGAGSDVDGDLGMAAGLQSLTTHDPSLQRYSEDPYVPLDSETDGYVAPLTC 1139
QY 1140 SPOPEYVNPQDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPGKGVVXDVAFAGAVEN 1199
DB 1140 SPOPEYVNPQDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPGKGVVXDVAFAGAVEN 1199

```

```

DB 1134 SPOPEYVNPQDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPGKGVVXDVAFAGAVEN 119
QY 1200 PEYLTPQGGAAPQCHPPAPASPAFDNLYYDODPPEGAPSPSTFKGTPRANEYGLDV 125
DB 1194 PEYLTPQGGAAPQCHPPAPASPAFDNLYYDODPPEGAPSPSTFKGTPRANEYGLDV 125
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
AC AAU77114;
XX 05-JUN-2002 (first entry)
DT
DE Human Her-2/neu polypeptide.
KW Human; Her-2/neu; cytostatic; hematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
OS Homo sapiens.
PN MO200213647-A2.
XX 21-FEB-2002.
PD
PE 13-AUG-2001; 2001MO-US25408.
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
PA (CORI-) CORIYA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI: 2002-280741/32.
DR N-PSDB: ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp: English.
XX
CC The invention relates to a method for inhibiting development of
CC hematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of hematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.1%; Score 6655; DB 23; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2;
QY 1 MELALACRGLLALLPFGAASVCTGTDKMLRLPAPBETHLMDLRLYOGCGVQGNL 60
DB 1 MELALACRGLLALLPFGAASVCTGTDKMLRLPAPBETHLMDLRLYOGCGVQGNL 60
QY 61 ELTYLPFTNASLSFIQDIOEVOGYVLIHNNQVRYPLQRLRIVRGTQLFEDNYALAVLDNG 120

```

```

Db      61 ELTYLPTNASTLFIQIOVGIVLIAHNOVRQPLQRLIRVRTQQLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Db      121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSPACHPSPCKGSRCKWGESSEDCOSLTRTVACAGGACRCKGPLPTDCCHQC 240
Db      181 LTLIDTNRSPACHPSPCKGSRCKWGESSEDCOSLTRTVACAGGACRCKGPLPTDCCHQC 240
Qy      241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCTYACP 300
Db      241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCTYACP 300
Qy      301 YNYLSTDVSGCTVCPILHNOEVTAEQNRCEKSKCAVCYGLGFENFTSPMLR-V 359
Db      301 YNYLSTDVSGCTVCPILHNOEVTAEQNRCEKSKCAVCYGLGFENFTSPMLR-V 359
Qy      360 PKVSASHLEEFAGCKKIFGSLAPLPESFDGDPASNTAPQYIKANSKFIQITETGYLIS 419
Db      360 PKVSASHLEEFAGCKKIFGSLAPLPESFDGDPASNTAPQYIKANSKFIQITETGYLIS 419
Qy      420 AMPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSLRPLSGLALIHNNTH 479
Db      420 AMPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSLRPLSGLALIHNNTH 479
Qy      480 LCFVHTVPMQDLFRNPQALHTANPEDECEVGEGLACHQLCARGHGWGPGPTQCVNCSQ 539
Db      480 LCFVHTVPMQDLFRNPQALHTANPEDECEVGEGLACHQLCARGHGWGPGPTQCVNCSQ 539
Qy      540 FLRGOECVEECRVIOGLPREYVNAHCLPCHPECOPOGNSVTCGPEADQVCAHAYKDP 599
Db      540 FLRGOECVEECRVIOGLPREYVNAHCLPCHPECOPOGNSVTCGPEADQVCAHAYKDP 599
Qy      593 PFCVACRCPGVPKPLSYMPIMKFPDEBGAOCPCINTHSCVDLDBKGCABEASAPLTS 659
Db      593 PFCVACRCPGVPKPLSYMPIMKFPDEBGAOCPCINTHSCVDLDBKGCABEASAPLTS 659
Qy      660 TVSAVVGILLVVLGVVFGILLIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAQR 719
Db      660 TVSAVVGILLVVLGVVFGILLIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAQR 719
Qy      720 ILKETELRKVULSGAGTGYKGIWIPDGENYKIPAIKVLKENSPPKANKILDEAYV 779
Db      720 ILKETELRKVULSGAGTGYKGIWIPDGENYKIPAIKVLKENSPPKANKILDEAYV 779
Qy      774 MAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLDHYRENRGRIGSODLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLDHYRENRGRIGSODLLNMCQIAKGM 839
Qy      840 SYLEDVRLVHARDLAARNVLYKSPNHVKITDGLARLIDIDETEHADGKVPIKMALES 899
Db      840 SYLEDVRLVHARDLAARNVLYKSPNHVKITDGLARLIDIDETEHADGKVPIKMALES 899
Qy      900 ILRRRFTHQSDVMGSGVTVMELMTFGAKPYDGIPIREIPDLLEKGERLPORPCTIDVYM 959
Db      900 ILRRRFTHQSDVMGSGVTVMELMTFGAKPYDGIPIREIPDLLEKGERLPORPCTIDVYM 959
Qy      960 IMVKCMIMIDSECRPFEBLVSEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1019
Db      960 IMVKCMIMIDSECRPFEBLVSEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1019
Qy      1074 FLAPSEGAGSDVFEDGDIQMGAAKGLQSLPTHDSPLQRYSEDDTVPLPSFTDGYAVPLTC 1133
Db      1074 FLAPSEGAGSDVFEDGDIQMGAAKGLQSLPTHDSPLQRYSEDDTVPLPSFTDGYAVPLTC 1133
Qy      1140 SPOPEYVNOQPVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVMDVFAFGAVEN 1199
Db      1140 SPOPEYVNOQPVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVMDVFAFGAVEN 1199

```

```

Db      1134 SPOPEYVNOQPVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVMDVFAFGAVEN 1199
Qy      1200 PEYLTPQGAAPQHPHPAPSPAFDNLVYMDQDPERGAPSTFFKGTPTAENEYVGLDV 125
Db      1194 PEYLTPQGAAPQHPHPAPSPAFDNLVYMDQDPERGAPSTFFKGTPTAENEYVGLDV 125
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein: 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WC-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
XX
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston TX, Huston JS, Oppermann H, Ring DB;
XX
DR WPI: 1993-272889/34.
XX
DR N-PDSB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX
PS antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX
CC tumour cells, such as breast and ovarian tumour cells, which is an
XX
CC approx 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
XX
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA:

Query Match 96.5%; Score 6616; DB 14; Length 1433;
Beet Local Similarity 96.5%; Pred. No. 0;
Matches 1223; Conservative 13; Mismatches 18; Indels 8; Gaps 2

Qy      1 MELALCRMGILLALLPAPGAASQVCTGTDMKRLPASBETHLDMRLHYOGQVVOGNTL 60
Db      1 MELALCRMGILLALLPAPGAASQVCTGTDMKRLPASBETHLDMRLHYOGQVVOGNTL 60
Qy      61 ELTYLPTNASTLFIQIOVGIVLIAHNOVRQPLQRLIRVRTQQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASTLFIQIOVGIVLIAHNOVRQPLQRLIRVRTQQLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Db      121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSPACHPSPCKGSRCKWGESSEDCOSLTRTVACAGGACRCKGPLPTDCCHQC 240
Db      181 LTLIDTNRSPACHPSPCKGSRCKWGESSEDCOSLTRTVACAGGACRCKGPLPTDCCHQC 240

```

Search completed: July 22, 2003, 08:41:29
Job time : 44.1589 secs

```
QY 241 AACCTGPRHSDDLACLFHNSGICELHCPALVTYNTDTPESMNPBGRYTFGASCTYACP 300
DB 241 AACCTGPRHSDDLACLFHNSGICELHCPALVTYNTDTPESMNPBGRYTFGASCTYACP 300
QY 301 YNYLSTDVSGCTLYCPLHNOEVTAEADGTQCEKSKPCARVCYGLGMFNNFTVSEWLR-V 359
DB 301 YNYLSTDVSGCTLYCPLHNOEVTAEADGTQCEKSKPCARVCYGLGMFNNFTVSEWLR-V 359
QY 360 PKVASHLEEFAGCKTFGSLAFLPESFGDPPASNTAPQYIKANSKFIGITELTGLYIS 419
DB 360 PKVASHLEEFAGCKTFGSLAFLPESFGDPPASNTAPQYIKANSKFIGITELTGLYIS 419
QY 420 AMPDLSPLDSVFQNLQYIRGRILHNGAVSLTLQGLISWGLRSRLRELSGLALIHNTH 479
DB 420 AMPDLSPLDSVFQNLQYIRGRILHNGAVSLTLQGLISWGLRSRLRELSGLALIHNTH 479
QY 474 LCFVHTVPMDOLEFNPQALLHTANRPEDCEVGEGLACHQLCARGHCMGPPTQCVCNSQ 539
DB 474 LCFVHTVPMDOLEFNPQALLHTANRPEDCEVGEGLACHQLCARGHCMGPPTQCVCNSQ 539
QY 540 FLRGOECVEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDP 599
DB 540 FLRGOECVEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDP 599
QY 594 PFCVARCPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTS 659
DB 594 PFCVARCPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTS 659
QY 660 IVSAVVGILLVVVLGVVFGILLIKRROKIRKYMRLLOETELVEPLTPSGAMPNQAQMR 719
DB 660 IVSAVVGILLVVVLGVVFGILLIKRROKIRKYMRLLOETELVEPLTPSGAMPNQAQMR 719
QY 720 ILKETELRKVKVLGSGAFGTVYKGIWIPGSENVKIPVAIKVLRNTPSPANKELLDEAVV 779
DB 720 ILKETELRKVKVLGSGAFGTVYKGIWIPGSENVKIPVAIKVLRNTPSPANKELLDEAVV 779
QY 774 MAGVSPYVSRLLGICLTSTVOLVTQLMRYGCLLDHYRENRGLSGQDLLNMCMQIAKGM 833
DB 774 MAGVSPYVSRLLGICLTSTVOLVTQLMRYGCLLDHYRENRGLSGQDLLNMCMQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKXMALES 899
DB 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKXMALES 899
QY 894 ILRRRFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIIDVYM 953
DB 894 ILRRRFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIIDVYM 953
QY 960 IMYKCMWIDSECRPRFELVSEFSRMAKRDQRFVVIQNEDLGPASPLDSTFYRSLIEDDD 1019
DB 960 IMYKCMWIDSECRPRFELVSEFSRMAKRDQRFVVIQNEDLGPASPLDSTFYRSLIEDDD 1019
QY 1014 MGDLDVAEEYLVPOQGFCDPPAPGAGMTHHRHRSSTRSGGDLTLGLPSESEAPRS 1073
DB 1014 MGDLDVAEEYLVPOQGFCDPPAPGAGMTHHRHRSSTRSGGDLTLGLPSESEAPRS 1073
QY 1080 PLAPSEAGSDVDVDGDLGMAAKGLOSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTC 1139
DB 1080 PLAPSEAGSDVDVDGDLGMAAKGLOSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTC 1139
QY 1134 SPOEYVNPQDVPRPQPSPREGFLPAPAPAGATLERPKTILSPGKGVKDVAFGAVEN 1199
DB 1134 SPOEYVNPQDVPRPQPSPREGFLPAPAPAGATLERPKTILSPGKGVKDVAFGAVEN 1199
QY 1200 PEYLTTPGGGAAPQHPHPPASPAFDNLVYWDODPBERGAPSTFKGTPTAENPEYLGLDV 1259
DB 1200 PEYLTTPGGGAAPQHPHPPASPAFDNLVYWDODPBERGAPSTFKGTPTAENPEYLGLDV 1259
QY 1260 PV 1261
DB 1254 PV 1255
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.1815 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-5-25-12

Perfect score: 6774

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGLDPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6672	98.5	1255	1	ERB2_HUMAN
2	5861.5	86.5	1254	1	ERB2_MESAU
3	5860	86.5	1257	1	ERB2_RAT
4	3133	46.3	1210	1	EGFR_HUMAN
5	3115	46.0	1210	1	EGFR_MOUSE
6	2968.5	43.8	1308	1	ERB4_HUMAN
7	2949	43.5	1308	1	ERB4_RAT
8	2696.5	39.8	1167	1	XMRK_XIPMA
9	2423	35.8	1342	1	ERB3_HUMAN
10	2356.5	34.8	1339	1	ERB3_RAT
11	1927	28.4	1426	1	EGFR_DROME
12	1749.5	25.8	1434	1	ERBB_DROME
13	1703	25.1	604	1	ERBB_AVIER
14	1630	24.1	540	1	ERBB_ALV
15	1586	23.4	703	1	EGFR_CHICK
16	1300.5	19.2	1323	1	L1232_CAEL
17	1142.5	16.9	245	1	ERB2_MOUSE
18	722	10.7	1363	1	ILPR_BRALA
19	698	10.3	1300	1	IRR_MOUSE
20	696.5	10.3	1382	1	INSR_HUMAN
21	694.5	10.3	1607	1	WPR_LYST
22	691	10.2	1372	1	INSR_MOUSE
23	690	10.2	1297	1	INSR_RAT
24	687.5	10.1	1477	1	HTK7_HYDAT
25	687.5	10.1	1300	1	IRR_CAVPO
26	683.5	9.7	1367	1	IGIR_HUMAN
27	654	9.7	1373	1	IGIR_MOUSE
28	644	9.5	1370	1	IGIR_RAT
29	640.5	9.0	2146	1	INSR_DROME
30	612.5	9.0	1390	1	INSR_AEDAE
31	610.5	8.9	987	1	ERB4_HUMAN
32	603	8.9	984	1	ERB1_CHICK
33	591.5	8.7	984	1	Q07494 gallus gall

34	589.5	8.7	977	1	EPB2_MOUSE	Q03145 mus musculus
35	588	8.7	1114	1	RET_HUMAN	P07949 homo sapien
36	584.5	8.6	976	1	EPB2_HUMAN	P29317 homo sapien
37	583.5	8.6	984	1	EPB1_RAT	P09758 rattus norv
38	582	8.6	987	1	EPB4_MOUSE	P54761 mus musculus
39	579	8.5	902	1	EPB8_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPB4_XENLA	Q91571 xenopus lae
42	573.5	8.5	1053	1	FAK1_CHICK	Q00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.4	757	1	HTK7_HYDAT	P53358 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculus

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p15setb2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 15).
GN ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Senda K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen F., Gray A.,
RA McGrath U., Seeburg P.H., Ullrich A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN (3)
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=295967;
RA Senda K., Kanata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN (4)
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993)
RN (5)
RP FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMPHIREGULIN.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIALLY).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 C RESIDUES (BY SIMILARITY).
 C -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 C POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 C OF 0.792; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 C ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 C -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 C
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL Outstation
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@sib-sib.ch).
 C
 CC EMBL: M11761; AAA35808.1; JOINED.
 DR EMBL: M11761; AAA35808.1; JOINED.
 DR EMBL: M11762; AAA35808.1; JOINED.
 DR EMBL: M11763; AAA35808.1; JOINED.
 DR EMBL: M11764; AAA35808.1; JOINED.
 DR EMBL: M11765; AAA35808.1; JOINED.
 DR EMBL: M11766; AAA35808.1; JOINED.
 DR EMBL: M11730; AAA75493.1; -
 DR EMBL: M12036; AAA35978.1; -
 DR EMBL: X03363; CAA37060.1; -
 DR PIR: A25481; A35481.
 DR PIR: A24571; A24571.
 DR HSP: P1362; IFGK.
 DR Genew: HGNC:3430; ERBB2.
 DR MIM: 164870; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00669; pkinase; 1.
 DR Pfam: PF00577; Furin-like; 1.
 DR Pfam: PF01030; Recept_L_domain; 2.
 DR Pfam: PF02757; YLP_2; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 POTENTIAL.
 FT TRANSMEM 653 675 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT BINDING 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 239 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 572 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I -> V.
 FT VARIANT 655 655 I -> V.
 FT VARIANT 655 655 I -> V.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT SEQID 1255 AA; 137909 MM; 3929PDA040CF962 CRC64.
 Query Match 98.5%; Score 6672; DB 1; Length 1255;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
 1 METAAACRWGILLALPRGAASSTOYCT-----QYIKANSKFIGITELLYOGGCGVQGNL 54
 1 METAAACRWGILLALPRGAASSTOYCTGDMKRLPASEPETHLDMRLHYOGGCGVQGNL 60
 55 ELTYLPYNASLPIODIQEVQGYLLTAHQVQVPLQRLRYRGTOGFEDNYALAVDNG 114
 61 ELTYLPYNASLPIODIQEVQGYLLTAHQVQVPLQRLRYRGTOGFEDNYALAVDNG 120
 115 DPUNTTPTVGTGSPGRLQLRSLEIKGVALIQRNPOLCYOPTILMKDIFKHNQLA 174
 121 DPUNTTPTVGTGSPGRLQLRSLEIKGVALIQRNPOLCYOPTILMKDIFKHNQLA 180
 175 LTLIDNRSRACHPCSCMCKGSPCKWSESDQSLTRYCAGGRCRGPPTCCHEOC 234
 181 LTLIDNRSRACHPCSCMCKGSPCKWSESDQSLTRYCAGGRCRGPPTCCHEOC 240
 235 AAGCTGPHSDCLALPHNSGICEHLCPALVTYNTDTFESMPNREGYRTGASCCTACP 234
 241 AAGCTGPHSDCLALPHNSGICEHLCPALVTYNTDTFESMPNREGYRTGASCCTACP 300
 295 YNYLSTDVGSCTLVCPILNQBVTAEADGTQCEKSKPCARVCYGLGMEHLREVRATVAN 354
 301 YNYLSTDVGSCTLVCPILNQBVTAEADGTQCEKSKPCARVCYGLGMEHLREVRATVAN 360
 355 IOEFGCKKIKGSLAFPESEDDPASNAPLOPELOVFTETETITGYLYISAMPDLP 414
 361 IOEFGCKKIKGSLAFPESEDDPASNAPLOPELOVFTETETITGYLYISAMPDLP 420
 415 DLSVQNLQVIRGRILNHGAVSLTLOGLISWGLRSLELQSSGLALLHNTHLCFVHY 474
 421 DLSVQNLQVIRGRILNHGAVSLTLOGLISWGLRSLELQSSGLALLHNTHLCFVHY 480
 475 PMDQFRPNHQAHLHTANPEDECVBGLACHOLCARGHCWPGPTQCVCNSQPLRQEC 534
 481 PMDQFRPNHQAHLHTANPEDECVBGLACHOLCARGHCWPGPTQCVCNSQPLRQEC 540
 535 VESCRVLOGLPREVYNARHCLPCPEQOPNGSVYCFGEAPQCVACHYDPPFCVAC 594
 541 VESCRVLOGLPREVYNARHCLPCPEQOPNGSVYCFGEAPQCVACHYDPPFCVAC 600
 595 PSGVPRDLSYPIKFPDEBACOPCFINCTHSVDLDDKSGPARQASPLTISVAVG 654
 601 PSGVPRDLSYPIKFPDEBACOPCFINCTHSVDLDDKSGPARQASPLTISVAVG 660
 655 ILLVVLGVGVGILIKRQOKIRKYTRRLQETELVEBLTPSGAMPQAMRILKTEL 714
 661 ILLVVLGVGVGILIKRQOKIRKYTRRLQETELVEBLTPSGAMPQAMRILKTEL 720

```

QY 715 RKVKVLSGAFGVYKGIWIPDGENVPIPAIKVLEBENTSPKANKEILDEAVVMGVGSP 774
DB 721 RKVKVLSGAFGVYKGIWIPDGENVPIPAIKVLEBENTSPKANKEILDEAVVMGVGSP 780
QY 775 YVSLIGICLTSTQVLTOLMPYGCILDHRENRGLSGSODLLNWMQJAKMSYLEDR 834
DB 781 YVSLIGICLTSTQVLTOLMPYGCILDHRENRGLSGSODLLNWMQJAKMSYLEDR 840
QY 835 LVHRDLAARNVLKSPNVAKITDFGLARLIDIDETEVYHADGQKVPKXMALESIRRRPT 894
DB 841 LVHRDLAARNVLKSPNVAKITDFGLARLIDIDETEVYHADGQKVPKXMALESIRRRPT 900
QY 895 HQSDVMSYGVTVWELMTFGAKPFDGIPARIPLDLKGERLPOPPICITDVTMIMWCKM 954
DB 901 HQSDVMSYGVTVWELMTFGAKPFDGIPARIPLDLKGERLPOPPICITDVTMIMWCKM 960
QY 955 IDSECRPRFRELVSFBRMARDQRFVVIQNEIDLGFPASPLDSFVYSILDEDDMDGLVDA 1014
DB 961 IDSECRPRFRELVSFBRMARDQRFVVIQNEIDLGFPASPLDSFVYSILDEDDMDGLVDA 1020
QY 1015 EEYLVPOQGFCDPAPAGAGVYVHRRSSSTRSGGDDTLGLEPSEEBAPRSPPLAPSEB 1074
DB 1021 EEYLVPOQGFCDPAPAGAGVYVHRRSSSTRSGGDDTLGLEPSEEBAPRSPPLAPSEB 1080
QY 1075 AGSDVDFGDLGMAKAGLQSLPTHDSPLOKVSSEDPVTLBSETOGYAPLTCSPQPEYV 1134
DB 1081 AGSDVDFGDLGMAKAGLQSLPTHDSPLOKVSSEDPVTLBSETOGYAPLTCSPQPEYV 1140
QY 1135 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVDVAFGAVENPEYLTPO 1194
DB 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVDVAFGAVENPEYLTPO 1200
QY 1195 GGAPQPHPPAPSPAFNDLYWDDOPPERGAPPSFKTKPTAENPEYLGDLDPV 1249
DB 1201 GGAPQPHPPAPSPAFNDLYWDDOPPERGAPPSFKTKPTAENPEYLGDLDPV 1255

RESULT 2
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60353;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
CN ERBB2 OR NEU
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxId=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ishikawa T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -|- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

```

```

CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16295; BAA03801.1; -
CC HSP: P11362; 1RCX
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_kinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_kinase.
CC InterPro: IPR004019; YLP motif.
CC Pfam: PF00069; kinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC Pfam: PF02757; YLP; 2.
CC ProDom: PD000001; Euk_kinase; 1.
CC SMART: SM00261; Fur_3.
CC SMART: SM00219; Tyrc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TIR; 1.
CC PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC KX Proto-oncogene; Disease mutation.
CC FT SIGNAL 1 21
CC FT CHAIN 1 21
CC FT DOMAIN 22 1254
CC FT TRANSMEM 22 652
CC FT TRANSMEM 653 675
CC FT DOMAIN 676 1254
CC FT DOMAIN 158 368
CC FT DOMAIN 472 644
CC FT DOMAIN 720 987
CC FT NP_BIND 726 734
CC FT BINDING 753 753
CC FT ACT_SITE 845 845
CC FT DISULFID 195 204
CC FT DISULFID 199 212
CC FT DISULFID 236 244
CC FT DISULFID 240 252
CC FT DISULFID 255 264
CC FT DISULFID 268 295
CC FT DISULFID 299 311
CC FT DISULFID 315 331
CC FT DISULFID 334 338
CC FT DISULFID 511 520
CC FT DISULFID 515 528
CC FT DISULFID 531 540
CC FT DISULFID 544 560
CC FT DISULFID 563 576
CC FT DISULFID 587 594
CC FT DISULFID 597 623
CC FT DISULFID 600 623
CC FT DISULFID 626 634
CC FT DISULFID 630 642
CC FT MOD_RES 630 642
CC FT MOD_RES 1139 1139
CC FT MOD_RES 1247 1247
CC FT CARBOHYD 68 68
CC FT CARBOHYD 125 125
CC FT CARBOHYD 187 187
CC FT CARBOHYD 259 259
CC FT CARBOHYD 330 330
CC FT CARBOHYD 571 571
CC FT CARBOHYD 629 629
CC FT VARIANT 658 658
CC FT VARIANT 659 659
CC SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2B51 CR664;

```

Query Match

86.5%; Score 5861.5; DB 1; Length 1254;

Best Local Similarity 86.1%; Pred. No. 76-305;
Matches 1081; Conservative 59; Mismatches 108; Indels 7; Gaps 2;

QY	1	MELALRCWGLLLLPAGASTOVCT-----QYIANKSFQIGITELLYQGGCVTVQSL	54
DB	1	MELAMCGWGLLLLPAGASTOVCTCTGDMKLRLPASPEITHLDIYRHLHYQGGCVTVQSL	60
QY	55	ELTYLPTNASTLSPFODIQEVQGYLLAHNOVROYQLRLRIVRGTOLEFEDNYALAVLNDG	114
DB	61	ELTYLPANATLSFQDIQEVQGYLLAHNOVRYHPLQRLRIVRGTOLEFEDNYALAVLNDK	120
QY	115	DPLNNTPTVTASFGGLRELOLRSLTILKGGVLIQNPOLCYQDTILMKDIFHKNOQLA	174
DB	121	DPLNNTPTVTATGTREGRELOLRSLTILKGGVLIQNPOLCYQDTIVLMKDVPRKNNQLA	180
QY	175	LTLIDITNSRACHPCSPKCGSRGMSSEPCOSLRTTVCAGGACAKGFLPTDCCHQC	234
DB	181	PVDIDITNSRACHPCAC	240
QY	235	AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDTEFESMNPREGRYTFGASCVTACP	294
DB	241	AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDTEFESMNPREGRYTFGASCVTACP	300
QY	295	YNYLSTDVSGCTLWCPLNNOEVTAEQTCCKSCKSCACVCGLGMELHGRALTSAN	354
DB	301	YNYLSTDVSGCTLWCPLNNOEVTAEQTCCKSCKSCACVCGLGMELHGRALTSAN	360
QY	355	IOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOEQLOVELEETLGYLYISAMPDULP	414
DB	361	IOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOEQLOVELEETLGYLYISAMPDULH	420
QY	415	DLVSFQNLQVIRGILNNGAVSLTQGLGISWGLRLRLSGGLALIHNNHLCVHTV	474
DB	421	DLVSFQNLQVIRGILNNGAVSLTQGLGISWGLRLRLSGGLALIHNNHLCVHTV	480
QY	475	PMQDLFRNPQALHTANRPEDECVCEGLACHOLCARGHMGSGPTQCVNCSOFLRGQC	534
DB	481	PMQDLFRNPQALHTANRPEDECVCEGLACHOLCARGHMGSGPTQCVNCSOFLRGQC	540
QY	535	VEECRVLOGLFREYVNAHCLPCHPECQPNQSVTCFGRPADQVCAAHYKDPFCVAC	594
DB	541	VKECRVWKGLFREYVNAHCLPCHPECQPNQSVTCFGRPADQVCAAHYKDPFCVAC	600
QY	595	PSGVKPDLSYMPKPFDEBGAQPCPCINCTHSCVDLDDMGCAEGRASPLSIVAVNG	654
DB	601	PSGVKPDLSYMPKPFDEBGAQPCPCINCTHSCVDLDDMGCAEGRASPLSIVAVNG	660
QY	655	ILLVVLGVVFGILLIKRQOKIRKRYMRLLQETELVEPLTPSGAMPNOAKRILKETEL	714
DB	661	ILLVVLGVVFGILLIKRQOKIRKRYMRLLQETELVEPLTPSGAMPNOAKRILKETEL	720
QY	715	RKVAVLGSAGATYKGIWIPDGENVKIPVAIVLENTSPKANKETILDEAYVMAGVSP	774
DB	721	RKVAVLGSAGATYKGIWIPDGENVKIPVAIVLENTSPKANKETILDEAYVMAGVSP	780
QY	775	YVSRLLGICLSTVQVLTQMLPYGCLLDHYRENRGLSGODLLNMCQIAKAGSYLEVR	834
DB	781	YVSRLLGICLSTVQVLTQMLPYGCLLDHYRENRGLSGODLLNMCQIAKAGSYLEVR	840
QY	835	LVHRDLAARNVLSVSPHVKITDGLRLDIDETEHADGCVPIKMMALSIILRRRT	894
DB	841	LVHRDLAARNVLSVSPHVKITDGLRLDIDETEHADGCVPIKMMALSIILRRRT	900
QY	895	HQDVMISYGVVLMELMFGAKPYDGIIPAREIPDLLEKGERLPQPTCTIDVYMIWCKM	954
DB	901	HQDVMISYGVVLMELMFGAKPYDGIIPAREIPDLLEKGERLPQPTCTIDVYMIWCKM	960
QY	955	IDSECRPRFELVSESRMARDQRFVIVNEDLGPASPLDSTFYSLLEDMDGLDVA	1014
DB	961	IDSECRPRFELVSESRMARDQRFVIVNEDLGPASPLDSTFYSLLEDMDGLDVA	1020
QY	1015	EEYLVPOQGFECPPAGAGAMVHHRSSSTRSGGDLTLGLEPEEERAPPSLAPSSG	1074
DB	1021	EEYLVPOQGFECPPAGAGAMVHHRSSSTRSGGDLTLGLEPEEERAPPSLAPSSG	108
QY	1075	AGSDVFPDGLGMAKALQSLPHRDSPLORYSDDTVPLPSTDGVAPLTSPPQEVY	113
DB	1081	AGSDVFPDGLGMAKALQSLPHRDSPLORYSDDTVPLPSTDGVAPLTSPPQEVY	114
QY	1135	NQPDVPRQPPSPREGPLPARPAGATLTERAKTSLPKNGVYKQVAFAGAVENPEYLTQ	119
DB	1141	NQPDVPRQPPSPREGPLPARPAGATLTERAKTSLPKNGVYKQVAFAGAVENPEYLTQ	120
QY	1195	GGAAPQHPPEPPAPSPAFDNLVYWDQPPERGAAPSPFTKGTAFNENPEYLGDPV	1249
DB	1201	GGASQCPH-PPALCPAFDNLVYWDQPPERGAAPSPFTKGTAFNENPEYLGDPV	1254
RESULT 3			
DB	ERB2_RAT	STANDARD; PRT; 1257 AA.	
QY	ERB2_RAT	01-JAN-1998 (Rel. 06, Created)	
DB	ERB2_RAT	15-DEC-1998 (Rel. 37, Last sequence update)	
QY	ERB2_RAT	15-JUN-2002 (Rel. 41, Last annotation update)	
DB	ERB2_RAT	Receptor Protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)	
QY	ERB2_RAT	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor	
DB	ERB2_RAT	receptor-related protein).	
QY	ERB2_RAT	ERB2 OR NEU.	
DB	ERB2_RAT	Rattus norvegicus (Rat).	
QY	ERB2_RAT	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
DB	ERB2_RAT	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
QY	ERB2_RAT	NCBI_TaxId=10116;	
DB	ERB2_RAT	SEQUENCE FROM N.A.	
QY	ERB2_RAT	TISSUE=Neuroblastoma;	
DB	ERB2_RAT	MEDLINE=6118662; PubMed=3945311;	
QY	ERB2_RAT	Bargmann C.I., Hung M.-C., Weinberg R.A.;	
DB	ERB2_RAT	"The neu oncogene encodes an epidermal growth factor receptor-related	
QY	ERB2_RAT	protein."	
DB	ERB2_RAT	Nature 319:226-230(1986).	
QY	ERB2_RAT	SEQUENCE OF 852-905 FROM N.A.	
DB	ERB2_RAT	TISSUE=Sciatic nerve;	
QY	ERB2_RAT	MEDLINE=9122560; PubMed=2025425;	
DB	ERB2_RAT	Lai C., Lemke G.;	
QY	ERB2_RAT	"An extended family of protein-tyrosine kinase genes differentially	
DB	ERB2_RAT	expressed in the vertebrate nervous system."	
QY	ERB2_RAT	Neuron 6:691-704(1991).	
DB	ERB2_RAT	STRUCTURE BY NMR OF 650-668.	
QY	ERB2_RAT	MEDLINE=9215518; PubMed=1346753;	
DB	ERB2_RAT	Gullikch W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D.,	
QY	ERB2_RAT	Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;	
DB	ERB2_RAT	"Three dimensional structure of the transmembrane region of the proto-	
QY	ERB2_RAT	oncogenic and oncogenic forms of the neu protein."	
DB	ERB2_RAT	EMBO J. 11:43-48(1992).	
QY	ERB2_RAT	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.	
DB	ERB2_RAT	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-	
QY	ERB2_RAT	ALPHA AND AMPHIREGULIN.	
DB	ERB2_RAT	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein	
QY	ERB2_RAT	tyrosine phosphate.	
DB	ERB2_RAT	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.	
QY	ERB2_RAT	THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.	
DB	ERB2_RAT	-1- SUBCELLULAR LOCATION: Type I membrane protein.	
QY	ERB2_RAT	-1- PMW: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE	
DB	ERB2_RAT	RESIDUES (BY SIMILARITY).	
QY	ERB2_RAT	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.	
DB	ERB2_RAT	-----	
QY	ERB2_RAT	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
DB	ERB2_RAT	between the Swiss Institute of Bioinformatics and the EMBL outstation.	
QY	ERB2_RAT	the European Bioinformatics Institute. There are no restrictions on its	
DB	ERB2_RAT	use by non-profit institutions as long as its content is in no way	
QY	ERB2_RAT	modified and this statement is not removed. Usage by and for commercial	
DB	ERB2_RAT	entities requires a license agreement (See http://www.isb.ch/announce/).	

QY 1133 YVNDPVPPEPSEEGPLPAAPAGATLRAKTLSPGKGVVYKDVFAFGAVENPEYLT 1192
 Db 1141 YVNSEVQPPPLPEEGPLPPEVAGATLPRKTLSPGKGVVYKDVFAFGAVENPEYLV 1200
 QY 1193 PGGGAAPQPPHPFAPSPAFDNLVYWDQDPPERGAPSPSTFGTPTAENPEYLGIDVPV 1249
 Db 1201 PREGTASPFPHPSPAFSPAFDNLVYWDQNSSEQGPPEPFGTGTATENPEYLGIDVPV 1257
 RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 ID EGFR_HUMAN
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
 AC Q00688; Q9B282; Q9H2C9; Q9H2C9; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCB1_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328112;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";
 RL Nature 309:418-425(1984).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=918811;
 RA Reiter J.U., Mahle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garita J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN (5)
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161733;
 RA Reiter J.U., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schehl Sinclair C., Pearisall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms";
 RL Genomics 71:1-20(2001).
 RN (6)
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.U., Threadgill D.W., Danielson A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Mahle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN (7)
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krutger M., Solarsky L.S., Weber M.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells";
 RL Science 224:843-848(1984).
 RN (8)
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN (9)
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Kallie B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN (10)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinschington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-396(1987).
 RN (11)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN (12)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN (13)
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN (14)
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA";
 RL Nature 309:270-273(1984).
 RN (15)
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombajagiam M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN (16)

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth factors 13:121-132(1996).
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p110 (shown here), 2/p60/
 CC truncated isoform/TGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 CC -----
 Query Match 46.3%; Score 3133; DB 1; Length 1210;
 Best Local Similarity 49.4%; Pred. No. 1,3e-159; Indels 114; Gaps 23;
 Matches 626; Conservative 178; Mismatches 348;
 QY 11 LLLALLPPGAA-STQVC-----TQYRANSKFTIGITELLYQGQGVQVQGNLELYLPT 61
 DB 14 LLAALCPASRALERKKVCCGCTSNKLTQLGTFFEDHFLST-QRMENNCEVLLGNLEIIVYGR 72

QY 62 NASLSFLQDIOEVGVYLLIAHQVQVFLQSLRIYRGTLQFEDNYALALVDNGPLNNTT 12
 DB 73 NYDLSFLKIQVAGVYLLALNTVERILEMLQIIRGMYYENSVALLAVLSNYD----- 12
 QY 122 PYTASPGGLRELQSLTEILKGGVLLIQNPOLCYQDTILMKDI FHKNNLALTLDITN 18
 DB 127 ----ANKTGLKELPMRNIOELIHGAVRFSNNPALCNVSIQMRDIVSSDFLSNMSDFQN 18
 QY 182 RSARCHPSGPMKSGSRGCGESSEDOGLTRVACAGCA-RKGLPLPDCHECCAGCTG 24
 DB 183 HUGSQCKDPSLCPNGSCWAGEENQKLTIKICAQQSGRGSGSPSCCHNCAAGCTG 24
 QY 241 PMSDCLACLHNSHGICELCPALVTYNTDTPFSMPEBGRYFGASCYACPNYLTST 30
 DB 243 PRESQVLCRKPRDQATKDCPPLMLYNPTTYQMDVNPBEKTFGATCYKKCRNRYVT 30
 QY 301 DVGSTLVCPPLHNOVTVLADGTQRCCKSKFCARVCYGLGHEHLREVAVTSANIQFAG 36
 DB 303 DHGSCVRAAGADSYEM-EEDEVKCKCKGECRCRCVNGDIGEFGKDSLISINATNIKHFKN 36
 QY 361 CKKIFGSLAFIPESFDGPASNTAPLQPEOLQVETLEITGYIYSAMPDPLDLSVQ 42
 DB 362 CTSISGDHLIPVARGSGSFHTPLPDLQSDIDIKYKEIIFLLIQAMPBNRTDLHAFE 42
 QY 421 NLQVIRGILHNGAVSLTLQGLISWGLRSLRELGSGLALIHNTILCFVHTVPMQDLF 48
 DB 422 NLEIRGRKQHGQSLAVSLVSLNTISGLRELKESIDGDIISGNKOLCVANTINMKLLE 48
 QY 481 RNPQALLHTANRPDEBVEGSLACHOLCARGHGWGSPQCVNCSQFLNGQCEVEGCRV 54
 DB 482 GTSGKTKIIRNGNSKACQVCHALCSGEGGGEPPDCVSCRNVSGRECVDCNLT 54
 QY 541 LQGLPREYVNAHCHLPCHPECCOPNGSVTCFGEPAQCVACAHYKDPFPCVARGSPGVK 60
 DB 542 LEGERREVENSEICIQHPECLPANMITCTGRPDNDICQAHYIDPHEVCKTCPSAVMG 60
 QY 601 DLSYPIKPFDEBACQPCPINCITHSQVDDDKCAEPASPLISIVAVG---ILL 65
 DB 602 ENNTL-VKXVADAHVCHLCPHNTCYGCTGSGLEGCPFNPKIP--SIATGMVALILL 65
 QY 658 VVVGAVNGGILIKRROOKIRKYTRRLLOELVELPITPGAMPNQOMILIKETELRY 71
 DB 659 VVALDIG---LPMRRHIVKRTIRLLQERBELVEPLTPGEBANQMLILILETEFKI 71
 QY 718 KYLSGAGFVYKGIWIPDGENYKIPVAIVLEBNTSPKANKEILDEAYVMAGVSPYVS 77
 DB 716 KYLSGAGFVYKGIWIPDGENYKIPVAIVLEBNTSPKANKEILDEAYVMAGVSPYVS 77
 QY 778 RLGLICLTSTVQLVLTQMLPFGCLLDHRENRKRLGSDLLNMCMQIAKMSYEDVRLYH 83
 DB 776 RLGLICLTSTVQLVLTQMLPFGCLLDHRENRKRLGSDLLNMCMQIAKMSYEDVRLYH 83
 QY 838 RDLAARNVLYKSPNHVKITDFGLARLIDIDETEYHADGKVPYIKMMALLESILRRRTTHS 89
 DB 836 RDLAARNVLYKSPNHVKITDFGLARLIDIDETEYHADGKVPYIKMMALLESILRRRTTHS 89
 QY 898 DVNSGYTVWELMTFFGAKPYDGIIPAREIPLLEKGERLPPPICTIDVYVIMVCKMTDS 95
 DB 896 DVNSGYTVWELMTFFGAKPYDGIIPAREIPLLEKGERLPPPICTIDVYVIMVCKMTDS 95
 QY 958 ECRPREFELVSEFSRMARDPQRFVAVIQ-NDLCPAPLSTPFRSLLEDODMDLYDADE 101
 DB 956 DSRPKRELLIIFSKMARDPQRLVAVIQGDRRMHLPSPTOSNFRALMDEEDMDVDADE 101
 QY 1017 YLVQGGFPCPDPAAGAGVHHRRHSSSTRSGGDUJTLGPESEEARSPPLAPSEGAG 107
 DB 1016 YLVQGGFPCPDPAAGAGVHHRRHSSSTRSGGDUJTLGPESEEARSPPLAPSEGAG 107
 QY 1077 SDVFDGDLGMAAGKIGISLPTHPSPLOVSEDPYPLPSET--GGVAPPLCSQPEPEV 113
 DB 1042 SN-NSVACIDRNGQSCPIKEDSFLQKISSPDTGALTEDSIDDTFL-----PVPEYI 109

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 19 19 C -> S (IN REF. 2).
 FT CONFLICT 539 539 C -> W (IN REF. 5).
 FT CONFLICT 591 591 L -> F (IN REF. 4).
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
 FT CONFLICT 1210 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64.
 SQ SEQUENCE

Query Match 46.0%; Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 1,2e-158;
 Matches 630; Conservative 170; Mismatches 355; Indels 118; Gaps 25;

QY 11 LLLALLPAGAA--STOYC-----TQYIKANSKFIGITELLYQGQGVQGNELTYLFT 61
 DB 14 LITALLCAAGALAEKRCVCGGTSNLTQLGTFEDHFLSL--ORMYNCEVVLGNLEITYVGR 72
 QY 62 NASLSFQDIOEVQGYVLIANOVQVPLQRLIVRSTQLFEDNVALAVLDNDPLANTT 121
 DB 73 NYDSLFLKTIQEVAGYVLIANTVERIPLLENLQIRGNALYENTYVALLISN----- 124
 QY 122 PVTGASPGGLRELORELTTELKSGVLIQENPOLCYODITLMDI----FKKNQALATL 177
 DB 125 --YGTNRTGLELRELMRWLQELTIGAVFSNNPILCNNDITQIMWDIVQYVFMSTMSYDL-- 180
 QY 178 IDNRSBACHPCSPMCKSGRSCWSESSDCOSLTRVACAGCA--RCKGPLYPTDCCHQCAA 236
 DB 181 --QSHPSCKPCDPSGSCNGSCWGGEGENCQKLTIKIICAQCSHRKCRGSRPSDCHNQCAA 238
 QY 237 GCTGPKASDCLACHFNHSGICELHCPALVTYNTDFESNPNDEGRYTTGASCVTACPN 296
 DB 239 GCTGPRESDCLVCCQCFQDEATCKDTCPLMLYNPTTYQMDVNVBEGKYSFGATCVKCPFN 298
 QY 297 YLSTDVSGCTLVCPLEHNOEYTAEDGTORCEKSGPCARVCGYGLGMEHREVAVYNSANIQ 356
 DB 299 YVVTDHGSCYRACGPDYEV--EEDGIRKCKCKCGFCKRVGNGIGIEFKDTLSINATNIK 357
 QY 357 EFGACKKIFGSLAPESFGDPASNTAPLOPQLOVFTLEETIGYLYISAMPDLPDL 416
 DB 358 HFKYCTAISGDLHLPLVAFKGDSEFTTRPPLDPRELELTKEITGFLLIQAMPDMDTL 417
 QY 417 SVFONLQVIRGRILHNHNGAVSLTLOGISWLGRLSRELGSGLALHNHTHLCFNYHYPM 476
 DB 418 HAFENLEIIRGRTQHOQFSLAVVGLNTLSLGRSLKEISDGVVITISGRNLCYANTINW 477
 QY 477 DQLEFNHQAHLHTANRPEDECVGBGLACHQLCARHCGPGFTQCVNSQFLRQCEVE 536
 DB 478 KTLFGTPOKTKIMNNAEKDCAVNVHVCNPLCSSGCGMPERDVCQVNSRRECEVE 537
 QY 537 ECRVLOGLPREYVNAHRCLECHPECQPNQSVYCFEPREDDCCYACAHYKDPFCVACRPS 596
 DB 538 KCNILBGEPRFEVNSECIQCHPECPCPQAMNITCTGRGPDNCIQCAHYDGHCHCTCPA 597
 QY 597 GVKPDLSTYPMKFPDEEGACQPCPINCTHSCVDLDKCGCPAEORASPLTSSVSAVGLL 656
 DB 598 GIMGENTTL--VMYAADANNVCHLCHANCTYGCAGPQLOGCEVWPSGPKPISLATIGVGL 656
 QY 657 LVVVLGVVFGI--LIKRRQKIKRYTYRRLLOETELVEPLTPSGAMPNQAKMLKETEELR 715
 DB 657 LFTIV--VALIGIGLFMRRIHVRKTLRLLOERELVEPLTPSGEAPNQAHRIILKETEELR 715
 QY 716 KVKVLGSGAGTGYKGIWIPDGENVXI PVALIKVLRNNTSPKANKELIDBAVYMAQVGSFY 775

DB 716 KIKVLGSAFGTGYKGLMIERGEVXKIPAIKELREATSPKANKELIDBAVYMAQVNDPH 775
 QY 776 VSRLLGICLTSTVQLVQLMPYGLDLHRENRGRASQQLWMCQIAGMSYLEDVRL 835
 DB 776 VCRLLGICLTSTVQLVQLMPYGLDLHRENRGRASQQLWMCQIAGMSYLEDVRL 835
 QY 836 VHRDLARNVLYKSPNHVKTIDTFCGLRLIDIDREYVADGKVPKIMMALESILRRFTH 895
 DB 836 VHRDLARNVLYKSPNHVKTIDTFCGLRLIDIDREYVADGKVPKIMMALESILRRFTH 895
 QY 896 QSDVMSYGVTVWEIMLTGKAPYDGIAPREIPLLEKGERLPQPPICITIDYMYIMVCMWI 955
 DB 896 QSDVMSYGVTVWEIMLTGKAPYDGIAPREIPLLEKGERLPQPPICITIDYMYIMVCMWI 955
 QY 956 DSECRPPREIVSEFSMARDDPORFVYIC--NEDLGRASPIDSIFRYSLLDDMDGLVVA 101
 DB 956 DADRREKREILIEFSKWARDPOKYLIOQDERMHLPSPTDSNFRALNDEBUEDEVDA 101
 QY 1015 EBYLVPOQGFPCDPAPAGAGMYHRRSSSTSGGDLTLGLEPSEEEAPRSPAPSEG 107
 DB 1016 DEYLIPQGGFF-----NSPST-----SRTPLLSLS 104
 QY 1075 AGSDVPFGDLGMAKGLQSLPTHDSPFLQYSEDPTVLPSET--DGYVAPLITCSPOPE 113
 DB 1042 ATSN--NSTVACINNGSCRYKEDAFIORYSDPTGAVTEBDNIDAFI-----FVPE 109
 QY 1133 YVNPDPVROPSPRESGLPARPAGATLERAKTLPGKNGVYKDVAFAGAVENPEYL- 119
 DB 1092 YVNO-SVPKRPAGVQNPVYHNPPLP-----APGRDLHQN--PSNNAVNEPYLN 114
 QY 1192 TPQGAAPQHPPPAPSPAFDNLVYWDQ-----DP-----PERGAPSTFKGTP 123
 DB 1141 TAQ-----PTCLSSGFPNSPALMIQKSHQMSLDNDPDYQDFFPKETKPGIFKG-P 119
 QY 1236 TAENPEYLG-DVP 1248
 DB 1191 TAENAEYLRVAPP 1203
 RESULT 6
 EMB4_HUMAN
 ID_EMB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor Protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112) (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERBB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=838332;
 RA Plowman G.D., Cullisou J.-W., Whitney G.S., Green J.M., Carlton G.W., Foy L., Neuhauer M.G., Shoyab M.;
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the Eidermal growth factor receptor family";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Kleenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D., Klagesrum M.;
 RT "A novel junctional domain isoform of HER4/Erbb4, isoform-specific tissue distribution and differential processing in response to phorbol ester";
 RL J. Biol. Chem. 272:26761-26768 (1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BEGACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIRGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID CEREBELLUM,
CC PITUITARY SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07868; AAB59446.1; -
DR HSSP: P1362; 1FGK.
DR Genew; HGNC:3432; ERBB4.
DR MIM: 600543; -
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR Prodom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 4.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 486 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPPLIC 626 648 NGPSSHCIYPPVGHSTLPQHA -> IGSSIEDICGLMD
FT (IN ISOFORM JM-B).
SQ SEQUENCE 1308 AA; 146807 MW; 5EABE0985D86761 CRC64;
Query Match 43.8%; Score 2968.5; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 8.4e-151;
Matches 609; Conservative 186; Mismatches 369; Indels 185; Gaps 30
9 WGLLALLPPGAA-----STGYCTGYIYANSKFGITEL-----LYGCGGVOCNLE 55
8 WYWSLVLAAGTVPDSQSYCA---GTENKLSLSLEQQYRALRYRNCVWGNLE 64
56 LTYPTNASTLFLODIOEVQGVYLAHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNGD 115
65 ITSIEHNHDLFFLSRVSEVGVYVALNQRYPYLEVRIIRKGLYEDRYALAFIYNR 124
116 PLNNTPTPTGASPGGLBELQRLSTTELLKGVLIQRPOLCYQDITLMKQIFPKNOAL 175
125 KDGNF-----GLQELGLKNTLETLLNGVYVDQKFLCYADTTHWQDIVENPMPNLS 175
176 TLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRIVCAGGC-ARCKGFLPTDCCHQC 234
176 TLVSINGSSGGGRCHKSGCTG-RCMGPENHCQITLITRVCAEQCDGRGVGVSDCHREC 234
235 AAGCTGPRGASCTCLAFHNSGTELHCPALVYNTDTFESMPNPRGRYTFGASCVYACP 294
235 AAGCGSPADTCCFCAMFNDSGACVTCQPTFYVNPPTFLPHNFNFKAYYGAFCVKKCG 294
295 YNYSSTDVGSCTVLCPLHNOVETADGTQCEKSKRCVAVCYGLGHEHREVAVTSAN 354
295 HNFV-VBSSSCVRAKPSKMEV-BENQIKCKRPTDLCPRACGIGISGLMSQYDSSN 352
355 IOEPACKKIGSLAFIPESFDGDPASNTAPLOEOJQVEFTELEITGYLYISAMPDSEP 414
353 IDKFNCTKIKINGNLIFFVTGIGHDPYVAIEAIDPEKINVRTRYREITGFANISWPPNMT 412
415 DLAVNONQVIRGRILHNGAVSLTLOGLGSMIGLRSLRELGGALIHNNTHLCVHNV 474
413 DSVFSNVLITGGVYVLSGLSLKLGQGLTSLQPSLKEISAGNIYITDNSVLCTYHTI 472
475 FMDOLFNPNGQALLATANRPEDEVCAGLAHQICARGHCWGPPTQCVNCSQFLRGQEC 534
473 NMTLFTSTINQRIYIDNRKAENCTAGCWCMHLCSSDGCWGPQDCSCRFSGRIC 532
535 VEEGRVLDGLPREVFNARHCLCPHEPCQ-PQNSVYTCFGEADQCACAHYKDPPTCVAR 593

```

Db      533  IESCNLYGSEEPREFENGSI CVECDPOCEKMEBDGLITGHPGPDNCTKCSHFKNPCNVEK 592
Qy      594  CPSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDDDDKGC-----PAEQRA 642
Db      593  CPDGLGANSF--IKRYADPRRECHPHCHPNCQCGNGFTSHDCIYIPWTGHTIPQAR- 649
Qy      643  SFLTSIVSAYV-GILLVYVLGVVFGILLKRRQOKTRKTYMRLLQETELVEPLTPSGAMP 701
Db      650  TPL--IAAGVIGGLILVYVGLTFVAVYVRKSIK-KKRALRRFL-ETELVEPLTPSGTAP 705
Qy      702  NQAOIRILKETELRKVKYLGSGAFGVKGIWIPGGEVVKIPVKIKYLRNTSKAKKEI 761
Db      706  NQAOIRILKETELRKVKYLGSGAFGVKGIWIPGGEVVKIPVKIKYLRNTSKAKKEI 765
Qy      762  LDEAYVAGVSGPVYSRLIGICLTSTVOLQTMPEYGCGLDHYENRGRGSDLLNWCN 821
Db      766  MDEALIMASMDPHLVRLIGVCLSTPTIQLVQTMHGCCLLEYVHEKDNIGSQLLINWCV 825
Qy      822  QIAKGMVSYLEDVRLVHRDLAARNVLSPNHVKITDPEGLAFLDIDSTEYHADGKYPIK 881
Db      826  QIAKGMVSYLEDVRLVHRDLAARNVLSPNHVKITDPEGLAFLDIDSTEYHADGKYPIK 885
Qy      882  WMALESILRRRTTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPIC 941
Db      886  WMALECIHYRKEFTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPIC 945
Qy      942  TIDVYMWVKCMIDSECRPRFRELSFSMSMADPQAFVITQNE-D-LGPASPLDSTFYR 1000
Db      946  TIDVYMWVKCMIDSECRPRFRELSFSMSMADPQAFVITQNE-D-LGPASPLDSTFYR 1005
Qy      1001  SLEDDMDGLVDASEYLVLPQGFCEPPAPAGAGMYHRRSSSTRSGGDLTLGLEPS 1060
Db      1006  NLDSEEDLEDMDASEYLVLP-QAFNIIPP-----ITSRKRLISNNS-----EIGHSP 1053
Qy      1061  BEAPRS-----PLAP-SEGAGSDVFPDGLGMAAGK 1091
Db      1054  PAYTWSGNQFYRDGFPAAQGVSVPYRAPSTIPAPVAQGTABIFDSCNGTLRK 1113
Qy      1092  LQSLPTHDPSPLQRYSEDPPTVPLPS-----ETDGVAPVLTCSPOPEYVNOGDVAPQP 1144
Db      1114  PVAPHVQDSSSTQKRSADPTVFAPERSRGRGLDEGIMTWRKPKQKOEYINPYE----- 1167
Qy      1145  SPREGPPLPARPAGATLERAKTISPKKNGVYKDVAFAGAVENPEYLLTPOGGAAPQPHP 1204
Db      1168  --ENPFVSR-----KNGDLQ-----ALDNEVYNAANG-----PP 1196
Qy      1205  PA-----FSPAFDNLVYVDQDPPPERGA--PP 1228
Db      1197  KAEDRYVNEPLVLTNPANTLGAELVKNLITSMPEKAKKAFDNPDYWNHSLPPRSTLQHP 1256
Qy      1229  STFKGTPT-----AENPEYL 1243
Db      1257  DYLOEYSTKIFYKONGRIPIVAINPEYL 1285

```

```

RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes.";
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      (2)
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system.";
RL      Neuron 6:691-704(1991).
RN      (3)
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Rohmert P.W., Kim S.S., Corbett J.A.;
RT      Expression of neuregulin and their putative receptors, Erb2 and
RT      Erb3, is induced during Wallerian degeneration.";
RL      J. Neurosci. 17:1642-1659(1997).
CC      - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      - SUBCELLULAR LOCATION: Type I membrane protein.
CC      - TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF041838; AAC08899.1;
CC      HSSP: P11362; IFCK.
DR      InterPro: IPR000494; EGFR_L domain.
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR002174; Furin-like.
DR      InterPro: IPR001245; Tyr_pkinase.
DR      InterPro: IPR004019; VLP motif.
DR      Pfam: PF00757; Furin-like; 1.
DR      Pfam: PF00068; Pkinase; 1.
DR      Pfam: PF01030; Recep_L_domain; 2.
DR      Pfam: PF02783; VLP; 2.
DR      PRINTS: PR00109; TYRKINASE.
DR      ProDom: PD000001; Euk_pkinase; 1.
DR      SMART: SMO0261; Fu; 4.
DR      SMART: SMO0219; TyrKc; 1.
DR      PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR      PROSITE: PS0011; PROTEIN KINASE_DOM; 1.
DR      PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
KW      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT      SIGNAL 1 25
FT      CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT      DOMAIN 26 651
FT      TRANSMEM 652 675
FT      DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP_BIND 718 985 PROTEIN KINASE.
FT BINDING 724 732 ATP (BY SIMILARITY).
FT ACT_SITE 731 731 ATP (BY SIMILARITY).
FT DISULFID 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D94HB096A08B41 CRCC4;

Query Match 43.5%; Score 2949; DB 1; Length 1308;
Best Local Similarity 45.0%; Pred. No. 9, 2e-150;
Matches 607; Conservative 193; Mismatches 378; Indels 172; Gaps 30;

1 MEUA-ALCEWGLL--ALLPFGASTOVCTQYIKANSKFIGITEL-----LYGCGQ 48
1 MKLATGLMWGSLLVAAFTVQPSASQVCA---GTENKLSLSLDEQYRALRKTYENCE 57
49 VVUGNLELYLPFNASLFLQDIQEVQVYLIAHQVQVPLQRLIVRGQLFEDNYAL 108
58 VVMGNLEITSIEHNRDLSEIRSVETGYVALAQFRYLPLENIRIRGKLYEDRYAL 117
109 AVLDNGDPLNNTTPVYGASPGG:REIOLSLTEILKGVLIQIBNRQLCYOQTLIMKFIH 168
118 AIFLNYRKQGNF-----GLQELGLKNLLEILANGVYVDQNFLELYADLIHMQDIVR 168
169 KNNQALLTIDINRSACHPGSPMCKGSRCEWSESSEDCQSLRTVYACAGC-ARCKGPLPT 227
169 NPWPSNMTLVSTIGSGGCRCHKSGCTG-RCMGPTENHCQTLRTVCACQCGRCYGPYVS 227
228 DCCHQCAAGCTGPKRSDCLACLFHNSGICELHPALVTYNTDFFESMPBEGRYFGA 287
228 DCCHRECAAGCGSPKDTDFACMNRNDSACVTCQCPQTFVNPFTFPLEHNFNAKYTGA 287
288 SCVTACPIVYISTDVSGCTLVCP:LNQEVTAEDGTQRCSEKSKPCARVCYGLGMEHLREV 347
288 FCVKKCKPHNFV-VDSSSCVRAKCPSSSKMEV-BENGIMKMKPCTDLCPRKACDSIGTGLMSA 345

QY 348 RAYTSANIOEFACCKIFGSLAFIPSPFGDASNTAP:QPEOLOVFETLEBITGYLIS 40
Db 346 QYDSSNIDDFINCTINGNLIFLYTGIGHDPPYNAIDADPEKLVNFRVTRBITGLNQ 40
QY 408 AMPDSLPLSVFONLOVIRGRILHNGAYSLUGLGISWGLRSIRELSGLLIHNNH 46
Db 406 TWPMNMTDFVFNVLITIGRVLVYSLSLILIKQOQITSLQFOSLKEIAGNVIYIDNSN 46
QY 468 LCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGILACHQLCARGHCMWPGPQVCNSQ 52
Db 466 LCYYHTIMNTTLTSTYNORIVIADNRAENCTAEGVNCNHLCSNCGMCPGPDQCLSCR 52
QY 528 FLRGQECVEECRVLQGLPREVYVARNCLCHECCQ-QNGSVTTCGPREDDQVACAHYD 58
Db 526 FSRGKICIESCNLYDEFFREFENGSLCVCDSDCEMDEGLTCGPGPDNCTKCSHFQD 58
QY 587 PFCVACRPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDDLDDKGC----- 63
Db 586 GPNCEKPCPDVLQGANSF--IFKYADQDRECHPCHNCTQCGNCPHSHDCIYYPMTHST 64
QY 637 -PAEQASPLTISVASV-GLLVVYLVGVPEGLIKRRQOKIRKXTMRLLQETLVEPL 69
Db 644 LPQPAR-TPL-IAAGVIGLFLVLMALTFAVYVRRSIR-KKQALRFL-ETLVLVPL 69
QY 695 TPFGAMPNQAOMRILKETELRKVYLGSGAFQYVKGWI PDGENVKI:PVAIKVLRNTS 75
Db 699 TPFGTAPNQAOLRIKETELRKVKVUGSAGFQYVKGWIWPGETVVKI:PVAIKIINETTG 75
QY 755 PRANKETLDEAYMAGVSPYVSRLLIGITLSTVOLVTQLMRYGCLLDHVRNRRGLSQ 81
Db 759 PRANVEFMDEALIMASVDHPHVRFLGVCLSPITQVLVTQLMRHGCLLEVEHKNQISQ 81
QY 815 DLNMMCMQIAKGSYLEDVRLVHRDIAANVYKSPNHYKIDPGLAALDDDEFERYAD 87
Db 819 LILNMCVQIAKGMVIEERLVRHDLAANVYKSPNHYKIDPGLAALDEDEERYAD 87
QY 875 GGVKPIKMALESILRRPFTQSDVSYGVTVMEITGAKPYDGI:PAEIRPDLLEKGR 93
Db 879 GGVKPIKMALECIHRYKFTQSDVSYGVTVMEITGAKPYDGI:PAEIRPDLLEKGR 93
QY 935 LPQPICTIDVYIMVWKCMNIDSECRPRELVSFSFMADPQRFVYIYONED-LGPASP 99
Db 939 LPQPICTIDVYIYVWKCMNIDADSRPFKEILAAEFSSMARDPQRYVLIQDDDRMKLPS 99
QY 994 LOSTFYRLLEDMDMDGLVDAAEYLVPOQGFPCDP-----A 10
Db 999 NDSKFPQNLDEDELEDMDADAEYLV-QANVIPPITYTSRTRIDNSSEIGHSPRAYT 10
QY 1031 PQAQKVVHHRHSSSTRSGGDLTGLBPSSEEAAPRPLAPSEAGSDVPDGLAMGA 10
Db 1058 PMSGQFVYQDGGFAITQGG--MPVPYATITSTIPEAVA--QGATAMFDDSCNGTLR 11
QY 1091 GLQSLPTHDPSPLQRYSEDPYVPLS-----ENDGVAPLUTCSPOREYVNOQDVNRP 11
Db 1113 KPVPRVHQDSSSTQRYSDPVTYFABERPRABLDDEGIMTWMDKPKQBYLANPYE----- 11
QY 1144 PSBRSPFLPAAPPAATLERAKTSLPGNNGVVKQVFAFGAVENPEYLLPQGAAPQPH 120
Db 1168 ---ENPFVSR-----KNQDLO-----ALNDPEVHASASG-----P 11
QY 1204 PPA-----FSPAFLDLYWDDQPPRGA--P 12
Db 1196 PRADEBYVMEPLYNTFTNALGNMEIYMNKSLSPBEAKAKAFNDPWNHSLPRLTLOH 124
QY 1228 PSTFKGTPT-----AENPEYL 1243
Db 1256 PDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
XMRK_XIIPMA
ID XMRK_XIIPMA
AC P13388;
STANDARD; PRT; 1167 AA.


```

Db      593 CPHGILGSDPTL-IMKXADKMGCCQCHQNCCTGCGGPGGLSGCRGP-1YSHSLAVGLVS 650
Qy      654 GILLVVLVSVFGLILKROQKTRKTYMRLLQETLVLPPLPSGAMPQAOMLIKETE 713
Db      651 GLITVYVALLIVLLRRRRIRK-RKRTIRCLQEKELVPLPPSQAPQOATRLIKETE 709
Qy      714 LRKVKVLGSGAGTGVYKGIWIPGGEVVKIPVAIKVLRNTSPKANKELIDEYVWAGVGS 773
Db      710 FKQDRVLGSGAFGVYKGLMNPGENIRIPVAIKVLRNTSPKANKELIDEYVWAGVGS 769
Qy      774 PYVSRLLGICLTSTVQVLTQLMYGGCLLVHRENRGLSGQDLNMCQIAKMSYLEDV 833
Db      770 PIVCRLLGICLTSAVQVLTQLMYGGCLLVHRENRGLSGQDLNMCQIAKMSYLEDV 829
Qy      834 RLVAHRLAARNVIVKSPNHKIDPGLARLDIDETEVADGKVPKIMMALESLTRRF 893
Db      830 HLVAHRLAARNVILKSPNHKIDPGLARLDIDETEVADGKVPKIMMALESLTRRF 889
Qy      894 THQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVYIMVWCM 953
Db      890 THQSDVWSYGVTVWELMTFGSKPYDGIAPAREIPDLLEKGERLPQPICTIEVYMIILKCM 949
Qy      954 MIDSECRPRFRELVSFESMARDPQGFVVIQONEDLGPASPLDSTYRSLDDEDMGLVD 1013
Db      950 MIDPSRPRFRELVSFESMARDPQGFVVIQONEDLGPASPLDSTYRSLDDEDMGLVD 1004
Qy      1014 ABEYLVPOQGFPCPDPAFGAGVHRRHRSSTRSGGDLTLGLEPSEEAAPRSAPSE 1073
Db      1005 ADEYLLPYKRI-----NRQSS-----ECCIPPT 1027
Qy      1074 GAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPV-PLPSTDGVAPLTCSPQPE 1132
Db      1028 GH-----PVRENSTLTLLNISDPQNALKXLDH-----E 1057
Qy      1133 YVNPDPVRPQ-----PSPRE-----GPLP-AARPAGATLERAKTUSPGKNGVWCVF 1179
Db      1058 YVNPQGESSTSRSLDLYNNYEDLTGCMGPVLSQGEATNSREYNTQNSL---PL 1114
Qy      1180 AFGAVENPEYLTPOGGAAPQHPHPAPAFADNLVYWDQDPERGAPPSFFKGTPTAEN 1239
Db      1115 VSSGSMDDDDY---QAG-----YQAAF-----LPQTGALTGMGMLPAEN 1152
Qy      1240 PEYLG 1244
Db      1153 LEYLG 1157

```

```

RA      P10man G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA      Todaro G.J., Shoyab M., and expression of an additional epidermal growth
RT      factor receptor-related gene".
RT      Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909 (1990).
RN      [3]
RP      TISSUE=Placenta;
RX      MEDLINE=9328282; PubMed=7685152;
RA      Kachon W., Yazaki Y., Sugimura T., Terada M.;
RT      "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT      tyrosine kinase.";
RL      Biochem. Biophys. Res. Commun. 192:1189-1197 (1993).
CC      -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC      (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC      SECRETED (SHORT FORM).
CC      -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC      SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC      -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC      SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC      AND PROMOTES ITS ASSOCIATION WITH THE P45 SUBUNIT OF
CC      PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC      -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaborati
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on i
CC      use by non-profit institutions as long as its content is in no w
CC      modified and this statement is not removed. Usage by and for commerci
CC      entities requires a license agreement (see http://www.isb-sib.ch/announc
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M29366; AAA35790.1; -
DR      EMBL; M4309; AAA35979.1; -
DR      EMBL; S61953; AAB26935.1; -
DR      PIR; A36223; A36223.
DR      HSSP; P11362; 1FGK.
DR      Genew; HGNC:3431; ERBB3.
DR      MIM; 190151; -
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_Pkinase.
DR      Pfam; PF0069; Pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      ProDom; PD000001; Euk_Pkinase; 1.
DR      SMART; SMC0261; FU_3_Pkinase; 1.
DR      SMART; SMC0219; TyKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR_FALSE_NEG.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW      Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW      Alternative splicing.
FT      SIGNAL 1..19
FT      CHAIN 20..1342
FT      DOMAIN 644..664
FT      TRANSMEM 644..664
FT      DOMAIN 665..1342
FT      DOMAIN 709..966
FT      NP_BIND 715..723
FT      BINDING 742..742
FT      ACT_SITE 934..934
FT      DISULFID 186..194
FT      DISULFID 190..202

```

```

RESULT 9
ERBB3 HUMAN STANDARD; PRT; 1342 AA.
ID ERBB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Iseling W., Miki T., Pinescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDLCMDTIDRDIYRDDADAIIVKDNKR
SC -> GQFVPSVGLTPROADOMYLLDDPRRLTLISASSK
VPTLTAIV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.8%; Score 2423; DB 1; Length 1342;
 Best Local Similarity 40.7%; Pred. No. 9,7e-122;
 Matches 532; Conservative 190; Mismatches 446; Indels 140; Gaps 33;

```

QY 5 ALCRWGLLALLPRGAASTVQYIYKANSKFGITELLVQGGQVYQGNLELYPTNAS 64
DB 27 AVCP-GTLNGLSVTDENQYQTLY-----KLYERCEVVMGNLEIVLTGHMAD 73
QY 65 LSFLODIOEVGYVLIANQVQVLOLRIVRGTOLEFENNVALALVDNDPLANTTPT 124
DB 74 LSFLOMIREVTGYVLAMNESTPLPLNLRVVGSTGYVDSKFAIFVM-----LNVNT--- 125
QY 125 GASPDGLRELDRLSLEILKGVLIQENPOLCYODTILMKDIFHKKNQALATLIDTNR 184
DB 126 -NSSHALRQLRLTGLTEILSGVYIEKNDKLCMDTIDRDIYRDD---AEIVKDNCR 181
QY 185 ACHFCSPKCKGRWSSSEDCSLRTVAGGC-ARCKGRLPTDCHEQCAAGCTGPRH 243
DB 182 SCPPEHEVCCKG-RWGGGSEDCQTLTKTICAPQNGHCFEPFNQCCHCECAGGSGPDD 240
QY 244 SDCLACLFHNSGICEHCPALVTYNTDFESMPNEGRFTFGASCVTACPYNYLSTDVG 303
DB 241 TDCFACHFNDSGACVRCRCPPLVYNKLTQLLEPNHTIKYQYGVVASCENHVF-VDOT 299
QY 304 SCTLVCPPLHNOETAEDEGTORCEKSPCARVYGLGMEHLEVRVRAVTSANIQEFAGCK 363
DB 300 SCVACPPDKMEVD-KGLKMKCEPCGGLCPKACEGSGG--SRFTVDSNIDGFVNCCK 356
QY 364 IFGSLAFLPESFDPPASNTAPLOPELOVFEETLEITGVLYSAMPDDLPLSLYONLQ 423
DB 357 ILGNLDFLITGLNDPDKHRIKIPALDPEKLVFRVREITGLVNIQSPMPMWHFVSFNSLT 416
QY 424 VIRGRILHNGAYS-LTLQGLISWGLSRLSRELSGSLAIHNNTHLCFVHTVPMQDLFRN 482
DB 417 TIGRSILYNRGFSLLIKGNLNVTLGFRSLKEISAGRIYISANRQICVHNSLMTKVLRG 476

```

```

QY 483 PHOALHTA-NRPEDECVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQECVEECRLV 541
DB 477 PTERLIDIKHRRPRDCVABSKVCODPLCSSGCKMGPBGQCLSRNNSRSGVCVTHCNFL 536
QY 542 QQLPREYVNAHCLIPCHPECOQPNQSVTCTGPEADQCVCANHYKDPFVCARCPSPGKPD 601
DB 537 NGEPREFAHEACSCSCHPECCPMEGTATCNQSGSDTCAQCAPHRDGPHCVSSCPHGLG- 595
QY 602 LSYMPIMWPEEGACCPPIINCNSVDLDDGCPREGA-----SPLTSIVAVNGIIL 657
DB 596 -AKGPITYRPVQNECPCHENCTQCKGPELDQGLVILIKHTLNTALVIAQ--L 652
QY 658 VVVLGVVFGILIKRQOKIR-KYMRRLLOETELVEFLTPSGAMPNQAQRIILKETELRK 716
DB 653 VVIFMVLGFTLYRGRRIQNKAMRRYELRGSIEFLDPS-EKANVLAIRFKETELRK 711
QY 717 VKVLGSAFGVYVIGVITPDGENVKIYALKVLPENTSPANKELIDEAVYMAVGVSPVY 776
DB 712 LKVLGSGVFGVHGVWIPGESIKIPVCIKVIEDSKGRQSFQAVVDHMLAIGSLDHAH 771
QY 777 SRLIGICTSTVQLVLTQLMPEYGLLDHVENRRLGSGDLNMCQIAKMSYLEDVRLV 836
DB 772 VRLIGLCRGSGLQVLTQYVLPGLSLDHYRQRLGALGQLLLNNGVQIAKMYLIEHGVY 831
QY 837 HRDLAARVLYKSPNHVKITDFGLARLLDITETRYADGSKVDIKMMALESILRRFTHQ 896
DB 832 HRNLAARVLYKSPQVAVDFGVADLLPPDDQLLYSEAKTPIKMAALESIFHKYTHQ 891
QY 897 SDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKBERLPORPITIDYVMVVKCMMD 956
DB 892 SDVMSYGVTVWELMTFGABEPYAGIRLAIEVDDLLEKBERLAQPOICTIDYVMVVKCMMD 951
QY 957 SECPRERELVSERSNARDPQRFVYIQLNEDLPA---SPLDSTFYRSLLEDMDGLVD 101
DB 952 ENIRPTKEILANETRRARDPPRLVLIKRES-GPGLAAGEPRGLNKKLEBELLEPDL 101
QY 1014 AEYLTVQGGFPCDPRAPGAGVNHHRNNSSTRSGGDLTLCLEP-SEEFARSPLABS 107
DB 1011 LDLDLEAED-----NLATTTGSAISLTVGTJLNRPGSQSLSPS 105
QY 1073 EGASDVDFDDLGMAKGLQSLPTHD-PSPLQRYSEDDFVPLP-----SENDGYYVA-- 112
DB 1052 SGY-MPMNQNLGSSCOESAVSSSERCPRVSLH-----PMRGLCLASSESGHVTGS 110
QY 1124 -----PLTCSPOPE-----YNNQPDVRPQPSPREG-----LPAA 115
DB 1105 EAELOEKVSMCRSRSRSPRPGDSAYHSQRHSLTPTVPLSPGLSEEDVGVYMPDT 116
QY 1155 RPAGATLERAKTILSP-GKNGVY-----KDVFAFGAVENPEVLTQGGAAAPQRPHPZAF 120
DB 1165 HLKGTPESSREGTILSSVGLSVLGTBEEDD-----EEVEYMNRRRHSPPHPPPS 121
QY 1208 SPAPDNLVYWD-----QDPPERGAPSPFTFGPTLAENPEYL 1243
DB 1216 SLEELGYEYMDVGSJLASLGSQSCPLHPVPIMPAGITTPBEDYVYM 1263

```

RESULT 10
 ID ERB3 RAT STANDARD; PRT; 1339 AA.
 AC 062739; 062955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (C-erbB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN [1]
 SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=852190;
 RX Hellyer N.J., Kim H.H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 RN
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroligins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).

Query Match	Best Local Similarity	Score	DB 1	Length	339
5	ALCRNGLLALLP	34.88	Score 2356.5	DB 1	Length 1339
27	AVCP-GTLNGLSVTDADNQVTL	41.18	Pred. No. 3.4e-116	Indels 161	Gaps 34
Matches	523	Conservative 169	Mismatches 419	Indels 161	Gaps 34
125	GASPGGLRELDRLSRLTEILKGVLLIQNPOLCYODTILMIDIFHNKNNQALTLIDTRSR	184			
126	NSSHALRLQKLFQTLTEILSGVYIEKNDLCHMDITDWDIDYRVR	181			
135	ACHPSPWCKSRKRWGSSSDCCSLTITVACGG-ARCKPRLTDCCHCCOAACTGPRH	243			
182	NCPCHVEVCKG-RCMGPGPDCCQILITKICAPQNGRCFGPNPNOCCCHCACGCGSGPD	240			
244	SDCLACLFHNSGICELCPALVTYNTDFPESMPNREGRTFGASCTACPYATLTDVG	303			
241	TDFCRGFRFNSGACVRCRCEPLVYNTQLQLEBNPHTKYQYGVVACSPHNFV	299			
304	SCTLVCPILHNDVTAEDGTORCEKSPCARVYGLAMEHLREVRATVANSIOEFAGCKX	363			
300	FCVACAPDKKAEVD-KGLKWCCEPCGGLCPKACGEGTSG--SRVCTVSSNIDGFFVNCIK	356			
344	IFGSLAFIPESFDDPSNAPRLQPELOVFEELLETGLYLSAMPDLSPLDSVQNIQ	423			
357	ILGNDFLITGLANDPKNKIPALDPELVNFRVRELTGTLNLIQSWPPHANSVSNLT	416			
424	VIRRIILHNGAYS-LTQGLGISWLGRLSRLDELGLALIHNTTHLCFYVATWPDOLFEN	482			
417	TIGRSILYNRGFSILIMKNLNTVSLGRSLKEISAGVYISANQOCYHNHSLWTFLLNG	476			
483	PHQALLHTA-NRPDECVGGGLACHQICAGHGWGPGFQCVNCSQFLGQDEVEECRLV	541			
477	PSERLDIKYDRPLGECILAGKVCDBPLCSSGGGWGPGQCLSCRYVSBGVVTHCNFL	536			
542	QGLPREVYNAHCLCPKPECOPQNGSVTCGPAPADOCVACAHYKDPFVACRPSGVKPD	601			
537	QGEPRFVHAHQCSCHPECLPMBEGSTCNCGSSDCAACAHFRDPRHCVNSCPHIDLG	595			
602	LSYPIKIPFDEBACQPCPINCTHSC--VDLDDKCGPRABQRPASPLTISVAVGILLIV	659			

RA Clifford R., Schubach T.;
RN Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RL
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Lively E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RL hormone binding and kinase domains.";
RN Cell 40:599-607(1985).
[4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=2093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RL Drosophila EGF receptor homolog transcripts.";
RN Cell 46:1091-1101(1986).
[5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RX MEDLINE=99102120; PubMed=2982502;
RA Leschkin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RL specification in wild-type, Ellipse, and null mutant Drosophila.";
RN Dev. Biol. 205:129-144(1999).
[6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Holt R.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fendell M.D., Zhang Q., Chen L.X.,
RA Brattdon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu P.V., Berman B.P., Bhandari D., Bozhakov S.,
RA Beeson K.Y., Beus P.V., Bernal B.P., Bhargava J., Brothier P.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotcher P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.U., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Haverson D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali B., Kalush V., Kapen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA LaRo X.P., Lei Y., Levitsky A.A., Li J.Z., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris C., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Paley J.M.,
RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirtkas R., Tector C., Turner R., Ventner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weisenbeck G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yah R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=DaekwanYeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Dworkin S.C., Vincent W.S. III, Bilodeau-Wentworth D.;

RT growth factor receptor";
 RL Nature 314:176-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RX ANALYSIS.
 RA MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/f1b alleles: implications for
 RL the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perlmutter N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the Drosophila
 RL EGF receptor";
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A TRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOSEROSA
 CC AND VENTRAL NEURODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF058754; AAC08536.1; -
 DR EMBL, AF058753; AAC08536.1; JOINED.
 DR EMBL, AF058754; AAC08536.1; -
 DR EMBL, AF058752; AAC08535.1; JOINED.
 DR EMBL, K03054; AAAS1462.1; -
 DR EMBL, K03417; AAAS1460.1; -
 DR EMBL, K03416; AAAS0965.1; -
 DR EMBL, K03418; AAAS1461.1; -
 DR EMBL, AF109077; AAD26134.1; -
 DR EMBL, AF109078; AAD26132.1; -
 DR EMBL, AF109082; AAD26132.1; JOINED.
 DR EMBL, AF109076; AAD26133.1; -
 DR EMBL, AF109084; AAD26133.1; JOINED.
 DR EMBL, AF109079; AAD26130.1; -
 DR EMBL, AF109081; AAD26130.1; JOINED.
 DR EMBL, AF109079; AAD26131.1; -
 DR EMBL, AF109083; AAD26131.1; JOINED.
 DR EMBL, AF109089; AAD26135.1; -
 DR EMBL, AE003454; AAF46732.1; -
 DR EMBL, X02293; CAA26157.1; -
 DR EMBL, X78920; CAA55523.1; -
 DR EMBL, X78918; CAA55521.1; -

DR EMBL, X78919; CAA55522.1; -
 DR FIR, A00640; GQFE.
 DR HSSP, F11362; IFGK.
 DR F1yase; EFGN0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk Pkinase; 1.
 DR SMART; SM00264; FU; 7.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.4%; Score 1927; DB 1; Length 1426;
 Best Local Similarity 32.9%; Pred. No. 2,7e-95;
 Matches 460; Conservative 179; Mismatches 421; Indels 338; Gaps 41
 QY 44 YQCGVQVQGNLELYLPT-NASLFLDIOEVOGYVLIANOVROYDLPQLRIVRTGTLF 102
 DB 126 YTNCTYDGNKLTMLPKNENLDFLONINEVGYILISHVDYKVFPELQIRGRTL 185
 QY 103 -----EDVYALAVDNDGDPINNTPTVPGASPGGLRELQSLLEIKGGVLIQRNPQLCV 157
 DB 186 SLVSEEEKYALFV-----TSKYTLIEDLDVLDVNGVGFNNYNLCH 229
 QY 158 QDTLMDIFKKNQLALTLDTRNSAPCHPCSPMKSGRCSSESDCGLSTRVCAG 217
 DB 230 MRLTQSEIIVNGTDAYNYDFTAPERECKCHESCTHG-CWGEGRNKCQKFSKLCSPQ 288
 QY 218 CA--RCKGPLPTDCHEQCAAGCTGPRGSDCLCLFHNSGICELHCPALVYNTPTFES 275
 DB 289 CAGRCCTGPKRRECCHLFCAGGCTGTPQKDCIACKNPFDEAVSKECPPRKYNPTTYVL 348
 QY 276 MPNPEGRYTFGASCVTAPYNYLSTDVSGCTIVCPHLNDEVTAEDGTORCEKSKPCARV 335
 DB 349 ETNPEGRYAVGATCVKECP-GHLIRDNGACVRSQPDQKMDKGE-----CVCNCGCPMT 402
 QY 336 CYGLGMEHLEVRVAVTSANIOEFAGCKKIGSLAFPESDG--DPASTNA-----PLQ 387
 DB 403 CPGVTVLH-----AGNIDSRNCTVIDGNRIIDQTFSGQDYVANYTWGPRYTPD 454
 QY 388 PEQLQFETLEITGYLYISAMPDLSFDLSFONLOVIRGRIIANGAY-SLTQGGIGSM 446
 DB 455 PERREVSTYKEITGYINIGTHPQPNLSYFNNLETIRQLQMEFMALAIKSLVS 514
 QY 447 LGLASRELSSGLALIHNNTHLCFHTVPMWDLFRNPQALLHTANRPEDECVSGSLACH 506
 DB 515 LEMNNLQISSGSVIOHNDLCYVNSIRFPAIQKPEQGVWVNNIIRADLCERKGTICS 574
 QY 507 QLCARHGCMGPGFPQCVNCSQFARGQCVVEECVVLQCLPREVYNAHGLCPHCECPQNG 566
 DB 575 DQCNEDGCKWAGAGTDQCLTCGNFNFNGTCLADGYSINAK--FDNRTCIKICPECR----- 628
 QY 567 SVTCGPEADQCAACHAYKDPFCVRCAP-----SGVK----- 599

```

Db      629  --TCAGAGADHCQECVHVRDQHCVCSECPKXKYNDRGVCRECHATCDGCTGPKDTTIGIGA 686
Qy      600  -----PDLSTYMPIMKF--PDEEGACQP-----619
Db      687  CTTGNLAIINDATVKGCLMDKDKPD-GY--FMEYVHPOQSLKLAGRAVGRKCHPL 743
Qy      620  CFI-----NCTH-----626
Db      744  CELCTNYGHEOVCSKCTHYKREOCETECBPADHYTDEQRECFQRHREPCNGCTGPGADD 803
Qy      627  --SC-----VDLDDSG-----CPAEOR-----ASPLTS-----647
Db      804  CKSCNFKLPDANEGPYVNSTMRCSTCKCLERHNYQYTAICPYCAASPPSSKITA 863
Qy      648  -----TSAVNGILLVAVGVFGILIKRROOKIRKTY--MRLLQETELVEPLTPS 697
Db      864  NIDVMNIFITGAVLVPTICILCVV--TYICRKQKAKKETVKTMLSCGEDSEPLRPS 921
Qy      698  GAMPNQAMRIKETELKRVKLVLSGAFGTYYKGIWIPDGENYKIPATVYLENTPSKA 757
Db      922  NIGANLCKLRIVKDELKKGVLGKAFGRYKGVWPEGENVKIPAIKELIKSTGAE 981
Qy      758  NKEILDEAYVAVGSPYVSRLLGICLTSTVQVLTQMPYGCCLLDHYRENRGRGLSGODLL 817
Db      982  SEEFLEAYIMASEEHVYVLLKTLAVCWSSQWMLITQLMPGCLLDYRNNRDKIGSKALL 1041
Qy      818  NMCNOIAGMSYLEVRLVHRDLAARVLYKSPHNVKITOPGLARLLDIDETVHADGK 877
Db      1042  NMSTQIAKMSYLEERKLVHRDLAARVLYQTPSLVKITDFGLAKLLSSNRYKKAQK 1101
Qy      878  VPIKMALESILRRRFTHQSDVMSYGVTVWELMFPGAKPYDGIIPAREIPDLLEKGERLPQ 937
Db      1102  MPKMLAECEIRNNVFTSKSDVMAFGVTIMELLTFGRPHENIPADIDPLIIVGKLEQ 1161
Qy      938  PPICTIDVYIMTVKCMWIDSECRPFELVSESRVARDPQRFVIONEDLG--PASPDL 995
Db      1162  PEICSLDIYCTLLSCWMLDAMRPFTFQKLTTFEAFARDGRYLAIPGKFTRLPA----1217
Qy      996  STFYSRLIEDD---DMGDVVAEAYLPPQGFPGPDPAAGGVHHRHSSSTRSGGD 1052
Db      1218  ---YTSQDEKLIRKLKAPTIDGSAIKKPDYIQPKAFRPS-----HRTDT-----1262
Qy      1053  LTLGLEPSEEBAP-----RSLPAFSEAGSDVFDG--DLGMAKAGLSLPTHPSP 1103
Db      1263  -----DEMPXLNRYCKDPSNKNSSGTDDERDSAREVGNLR-----1300
Qy      1104  QRSEDEPTVPLPSTDGVVAPLTSPQPEYVQPDVAPQPSRREGPLPAAPDAGTLER 1163
Db      1301  -----LDLPVEDDYLMP--TCQPGPNNNNNM-----NENQNNMAVGVAAQYM--1343
Qy      1164  AKLSPKNGVAVKDFVAFGAVENPEYL---TPQGAAPQPH-----1202
Db      1344  -----DLIGVPSVNDPEYLLNAQLGIGESIPITQITIGIVMGPGTMEVK 1390
Qy      1203  -PPAFSP-AFDNLYYND 1218
Db      1391  VPMGSEPTSSDHEYND 1408

```

```

RESULT 12
ID      ERBB_ALV          STANDARD;      PRT;      634 AA.
AC      P00534;
DT      21-JUL-1986 (Rel. 01, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
OS      Avian leukosis virus.
OC      Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
NC      NCBI_TaxID=11864;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RX      MEDLINE=85228222; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Critchenden L.B., Raines M.A., Kung H.-J.,
RT      "c-erbB activation in ALV-induced erythroidleukosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-truncated EGF receptor.";
RL      Cell 41:719-726(1985).
CC      -| CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -| MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -| MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -| SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M10066; AAA48763.1; ALT_INIT.
CC      PIR: A00643; TVCHLV.
CC      PIR: B00643; TVPVLV.
CC      HSSP: P11362; IFGX.
CC      InterPro: IPR000719; Euk_Pkinase.
CC      InterPro: IPR001245; Tyr_Pkinase.
CC      Pfam: PF00069; Pkinase; 1.
CC      PRINTS: PR00109; TYRKINASE.
CC      ProDom: PD000001; Euk_Pkinase; 1.
CC      SMART: SM00219; Ty-KC; 1.
CC      PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC      PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC      TRANSFERASE: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      KW      Glycoprotein; Phosphorylation.
CC      FT      DOMAIN 132..399
CC      FT      NP_BIND 138..146
CC      FT      BINDING 165..165
CC      FT      ACT_SITE 257
CC      FT      BINDING 165
CC      FT      ACT_SITE 257
CC      SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

Query Match 25.8%; Score 1749.5; DB 1; Length 634;

Best Local Similarity 52.3%; Pred. No. 3e-86; Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

```

Qy      581  CAHYKDPFCVVARCPGKVDLSYMPIMKFPDEEGACQPCPINTCTHSCVLDLDCGCPAQ 640
Db      3  CAHFIDGPHCVKACPAVAGENDTL-VMKYADANAVCQCHPQCTRGCGPGLEGCP---58
Qy      641  RASPLTSIVSAVY-GILLVAVGVFGILIKRROOKIRKTYRMRLLQETELVEPLTPSCA 699
Db      59  NGKTFPIAGGVGGLLVVGLGVLRLRR-HIVRRRLRLRLQERELEVPPLPSE 117
Qy      700  MPNQAMRIKETELKRVKLVLSGAFGTYYKGIWIPDGENYKIPATVYLENTPSKANK 759
Db      118  APQAHRLIKETEPKVKVLSGAFGTYYKGIWIPDGENYKIPATVYLENTPSKANK 177
Qy      760  EILDEAYVAVGSPYVSRLLGICLTSTVQVLTQMPYGCCLLDHYRENRGRGLSGODLLNW 819
Db      178  EILDEAYVAVGSPYVSRLLGICLTSTVQVLTQMPYGCCLLDYRREKNDIGSQYLLNW 237
Qy      820  CMOIAGMSYLEVRLVHRDLAARVLYKSPHNVKITOPGLARLLDIDETVHADGKVP 879
Db      238  CVOIAGMSYLEERKLVHRDLAARVLYQTPSLVKITDFGLAKLLGAEKRYHAGGKVP 297
Qy      880  IKMMALESILRRRFTHQSDVMSYGVTVWELMFGAKPYDGIIPAREIPDLLEKGERLPQ 939

```

```

Db      298  IKMALESLHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIASISVLEKGERLPPQPP 357
Qy      940  ICTIDVYMIWYKMWIDSECRPFRELVSEFSRMARDQRFVVIQ-NEDLGPASPLDSTF 998
Db      358  ICTIDVYMIWYKMWIDADSRPFRELVSEFSKMRADPRYLVIQGDERRMLPSPDTSKF 417
Qy      999  YRSLEDDDMGDLVDAEYLVLPQCGFFCPDPAPAGAGWVHRHRSSTRSGGDLTLGLE 1058
Db      418  YRTLMEEDMEDIVDADEYLVHPQGF-----NSPST----- 449
Qy      1059  PSEEARPRSP-----APSEGAGSDVFQDGLMGAKLGSLPTHDSPFLORYSEDPVTP 1113
Db      450  -----SRTPLLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVORYSSDPTGN 495
Qy      1114  LPSET--DGVVAPLTCSPQPEVYNQDPVROPSPREGPLPARPAGATLERAKTLSPGK 1171
Db      496  FLEESIDDFL-----PAPRYVNO--LMPKPS-----TAVVQ 526
Qy      1172  NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDMLY 1216
Db      527  NQIYNNISLTALSKLPMDSRQNSHSTAVDNPEYL-----NTNQSPPLAKTVFESSPY 578
Qy      1217  WDQ-----DPEE-----RGAPSPSTFKGTPTAENPEYIGLDVP 1248
Db      579  WIDSGNHQINLNDPDIYQDPFLNETKPNGLKVPAPENPEYLRVAP 625

RESULT 13
ERBB AVIER STANDARD: PRT: 604 AA.
ID ERBB AVIER
AC R00515;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
   gene family.";
RL Cell 35:71-78(1983).
RN (2)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benajissa M., Bisette G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
   new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
   tyrosine phosphate.
CC -1- DISEASE: THE v-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
   ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
   IN CHICKENS.
CC -1- MISCELLANEOUS: v-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
   RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TYUH.
DR HSSP: P11362; IFGR.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk.pkinase; 1.
DR SMART: SM00219; TyrKc_1.
DR PROSITE: PS00107; PROTEIN KINASE, ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE, TYR; 1.
DR PROSITE: PS00111; PROTEIN KINASE, DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
   Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146
FT BINDING 165 165
FT ACT SITE 257 257
FT CONFLICT 29 29
FT CONFLICT 140 140
FT CONFLICT 146 146
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 25.1%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8,6e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 1

Qy 581 CAHYKDPFCVAPRCSPGVKPDLSYVPIKWPDEGACQPCINCTHSCVDLDDGCGAEQ 641
Db 3 CAHFDGPHCVACAGVUGENDTL-VAKIDANNAVCOLCPNCTRGKGGLEGP--- 58
Qy 641 RASPLTIVSAVV-GILLVVLGVVFGILIKRROQKIRKYMRLLOETELVEPLTPSGA 69
Db 59 NSKTPSIAGAVGGLICLVVGLDGLYLRR-HYAKRTLRLQRELVETPSGE 11
Qy 700 MNQAGMLKTEELKRYKVLGSGAFGVYVGINPDGENKIPVAILKVLRENTPRANK 75
Db 118 APNQHRLRLKTEKRVKVLGSGAFGIYVGLMPEGEKXIVAIKELREASPANK 17
Qy 760 EILDEAVYAGVGSPPVSRLLGICLTSTVQVLTQIMPGCLLDHVRNRCGSGQDLNW 81
Db 178 EILDEAVYASVDNHNVCRLGICLTSTVQVLTQIMPGCLLDHVRNRCGSGQDLNW 23
Qy 820 CNOIAGKSYLEDVLRDLAARNVLYKSNHYKTFDGLARLLIDETERYHAGGKVP 87
Db 238 CVOIKGMVYLEERLVRDLAARNVLYKTPQHYKTFDGLAKLIDEXEYHAGGKVP 29
Qy 880 IKMALESLHRRFTHQSVDVMSYGVTVWELMTFGSKPYDGIAPAREIDLEKGERLPPQ 93
Db 298 IKMALESLHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIAPASEISSVLEKGERLPPQ 35
Qy 940 ICTIDVYMIWYKMWIDSECRPFRELVSEFSRMARDQRFVVIQ-NEDLGPASPLDSTF 99
Db 358 ICTIDVYMIWYKMWIDADSRPFRELVSEFSKMRADPRYLVIQGDERRMLPSPDTSKF 41
Qy 999 YRSLEDDDMGDLVDAEYLVLPQCGFFCPDPAPAGAGWVHRHRSSTRSGGDLTLGLE 105
Db 418 YRTLMEEDMEDIVDADEYLVHPQGF-----NSPST----- 44
Qy 1059 PSEEARPRSP-----APSEGAGSDVFQDGLMGAKLGSLPTHDSPFLORYSEDPVTP 111
Db 450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVORYSSDPTGN 495
Qy 1114 LPSET--DGVVAPLTCSPQPEVYNQDPVROPSPREGPLPARPAGATLERAKTLSPGK 117
Db 496 FLEESIDDFL-----PAPRYVNO--LMPKPS-----TAVVQ 524
Qy 1172 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDMLY 121
Db 525 --VQIYNNISLTALSKLPMDSRQNSHSTAVDNPEYL-----NTNQSPPLAKTVFESSPY 574
Qy 1213 NLYYWDQDPFERGAPSPSTFKGTPTAENPEY 1242

```

```

Db      575 SSPWIOGSHQ-----INLNDPY 594

RESULT 14
ERBB_AVIEU STANDARD; PRT; 540 AA.
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064459; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -I- tyrosine phosphate.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13179; AAA42401.1; -.
CC PIR: A25231; TVEVER.
CC HSSP: P1362; IFGK.
CC InterPro: IPR000719; Euk_kinase.
CC InterPro: IPR001245; Tyr_kinase.
CC Pfam: PF00069; Pkinase; 1.
CC ProDom: PD000001; Euk_kinase; 1.
CC SMART: SMO0219; Tyrc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC KX
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H->D (IN THERMOLABILE V-ERBB).
CC SEQUENCE 540 AA; 60412 MW; 5853297AA068B65D CRC64;
SQ

Query Match 24.1%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 5.8e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 581 CAHYKDPFCVAPRPSGVKPLSLMPYMKRPDPSGACQPPINCTMSQVLDKGPAPQ 640
DB 3 CAHFDGPHCVKACPAVGLGENDTL-VKRYADANAVALOLHPNCTGCKGPGLEGPP--- 58
QY 641 RASPLTISVAVV-GILLVVLGVVFGILLIKRROKIRKTYRMRLLOETFLVPLTPSGA 699
DB 59 NGSKTPSIAAGVVGGLGLVVGIGLYLRR-HYRKRTLRLLQERLVEPLTPSGE 117
QY 700 MPNOQMRILKETLRKRYVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 759
DB 118 APNOAHRLKETLRFKRYKVLGFGAPGTIVKGLMIPGEKVTIPVAIKELREKTSKANK 177
QY 760 EILDEAVYMGVGSPPYVSRLLIGICTSTVQLVQMLQMYGGLLDHVENGRGLSQILLW 819

```

```

Db      178 EILDEAVYMGVSNPNHVCRLGLICLTSTVQLTQLMPFYCILLDIREHKONIGSQILLW 23
QY 820 CMQIAKGSYLEDRPLVYRDIAARNVLYKSNHVKITDFGLARLLDDEHYHADGKVP 87
DB 238 CVQIAKGMNVLSEERMYARDIAARNVLYKTPQHKITDFGLAKQLGADKELYAEGKVP 29
QY 880 IKMALSESILRRFTHQSDVMSYGVATWELMTFGAKPYDGIPIAEIPDLLEKGRLPQPP 93
DB 298 IKMALSESILRRITTHQSDVMSYGVATWELMTFGSKRYDGIPIAEISVYLEKGRLPQPP 35
QY 940 ICTIDVYIMYKCMWIDSECRPRREIVSEFSRMAPDPQRFVVIQ-NEDIGPASPDLSTF 99
DB 358 ICTIDVYIMYKCMWIDSECRPRREIVSEFSRMAPDPQRFVVIQ-OCGERMHLSPSTSKF 41
QY 999 YRSLEDDMDGLDVAEYVLPQGFPCPDPAEAGGMVHRRHSSSTRGGGLTLTGLE 10
DB 418 YRTLMEEEDMDIVADADYLVPHGFF-----NSPST----- 44
QY 1059 PSEEPAPRSPV-----APSEAGSDVDFGDLGMAKAGLSLPTHDSPLOQRYSEDTVP 11
DB 450 -----SRPPLSLSLATSNNSATNCIDRNG-----H----- 47
QY 1114 LPSEIDGYVAALTSPPQPEYVNOQDVNPQPPSPREGPLPAAPRAGAT-LERAKTISPKGN 11
DB 477 -PVREDGFL-----PAPEYVNO--LMPKRPSTAMVQNIYVISTLAISKLPIDSKYQN 52
QY 1173 GVYKDVFAFGAVENPEYL 1191
DB 528 -----SHRAVNVNPEYL 539

RESULT 15
EGFR_CHICK STANDARD; PRT; 703 AA.
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
LA lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
LA Ulrich A., Vennström B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988);
CC -I- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce

```

CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M20386; AAA48760.1; .
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR SMART; SMO0261; FU; 4.
 DR PROSITE; PS0107; PROTEIN_KINASE_ATP; PARTIAL.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT STGNL 1 30
 FT CHAIN 1 30
 FT DOMAIN 31 654
 FT TRANSMEM 31 654
 FT DOMAIN 655 667
 FT DISULFID 197 206
 FT DISULFID 201 214
 FT DISULFID 222 230
 FT DISULFID 226 238
 FT DISULFID 239 247
 FT DISULFID 243 255
 FT DISULFID 258 267
 FT DISULFID 271 298
 FT DISULFID 302 314
 FT DISULFID 318 333
 FT DISULFID 336 340
 FT DISULFID 313 322
 FT DISULFID 517 530
 FT DISULFID 533 542
 FT DISULFID 546 562
 FT DISULFID 565 581
 FT DISULFID 569 589
 FT DISULFID 592 601
 FT DISULFID 605 627
 FT DISULFID 630 638
 FT DISULFID 634 646
 FT CARBOHYD 134 134
 FT CARBOHYD 190 190
 FT CARBOHYD 200 200
 FT CARBOHYD 359 359
 FT CARBOHYD 368 368
 FT CARBOHYD 420 420
 FT CARBOHYD 573 573
 FT CARBOHYD 578 578
 FT CARBOHYD 613 613
 FT CARBOHYD 633 633
 FT CARBOHYD 648 648
 FT NON_TER 703 703
 SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.4%; Score 1586; DB 1; Length 703;
 Best Local Similarity 44.1%; Pred. No. 1.7e-77;
 Matches 312; Conservative 113; Mismatches 247; Indels 36; Gaps 14;

QY 8 RMGLLALLPPGAA-----STQVC-----TOYIKANSKFITGTELLLYOGCOVOGNTL 54
 DB 13 RGAALVLLLLGLGALCSAVEKVKCOGTNNKLTOLGHVEDHFTSL-QRMVNNCEVVLSTNL 71
 QY 55 ELTYLPTNASLSFLQDIEVQGVLIANQOVROVPLQRLIYRGTOLEPENVYALAVDNG 114
 DB 72 EITVEHNRDLTFKTIQEVAGVLIALNMWDVIPLENLQIIRGVNLYDNSFALAVLSNY 131
 QY 115 DELANTTPVTGASPGSLRFLQRLSLTEILKGVLIQRNPQLCYODTIIMKDIIFKNNOLA 174
 DB 132 H-MNKTQ-----GLRELPMKRLSEILINGVKISSNPKLCNMDVLMNDIIDSRRK-P 181
 QY 175 LTLID-TNRSRACHPSPCKSGRCMGSSSDCOSLTRVCAGGA-RCKGPLFTDCHE 232

DB 182 LTVLDPSASVLSGPCPKCHPCTEDHDCMGAGEONCQTLTVKICAGQCSGRGRGVPSDCCHN 24
 QY 233 QCAAGCTGKSHSDCLACHFNHSGICEIHCPLVYNTDTESMNPREGRYTFGASCYTA 29
 DB 242 QCAAGCTGRESDDCLACRFBDATCKDTCPLVLYNPTTYQMDVNPBGKYSFGATCYRE 30
 QY 293 CPNNYSTDVSGCTLVCPILHNOEVAIEAGTQRCCEKSCFPCARVVCYGLMEHLREAVATS 35
 DB 302 CPNNYVTHGSCVRSQNTDTEYV-EENGVRKCKKCDGLCSKVCNGIGIGELKGLISTNA 36
 QY 353 ANIQEPAGCKKIFGSLAFIPESFDDPASNTPALQPEQLQVETLEETGYLYISAMPDS 41
 DB 361 TNDSFKNCTKINGDVSLPVAFLGDAFKTLPLDPKXLDVFRVKEISGFLIQAMDN 42
 QY 413 LPLSLFQNLQVIRGRILHNGAYSITLQGLGISMGLSLSELSGLALIHNNHLCVH 47
 DB 421 ATDLVFEENLEIIRGRTOHQGYSLAVNLTQSLGLSLKEISDGLAIKNNKVLGYAD 48
 QY 473 TVPMDLFENPQALHTANREDECEVSGGLACHQLCARGCHWGPGPQVCVCSQFLRGQ 53
 DB 481 TMMWRSLFATQSQKTKIIONRKNKCTADRHVCDPLCSDVCGMGGRPHCSFRFSK 54
 QY 533 ECVBECRVLOGLPREYVNAHCLPCHPCQPNG---SVTCFGRPADQCAAHYKDPF 58
 DB 541 ECVKQCNIIQGEPRFERDSKCLPCHSECLVONSTAVNTCSGPGPDHCKCAHFIDGP 60
 QY 590 CVARCPGVPPLSLYMPIMKPPDEGACOPINQTHSCVLDLCKGCPAERASPLTSIV 64
 DB 601 CVKACPAGVLSGENDTL-VKXADANAVALQHPNCTRGCKGPGLEGCP--NGSKTPSLA 65
 QY 650 SAVV-GILLVVVLGVVFGILIKRQOKIRKXTMRLLQETELVEPLTP 696
 DB 657 AGVGGILCLVVVGIGIGVLR--HIVKRTLRLLQERELVEPLTP 703

Search completed: July 22, 2003, 08:43:43
 Job time : 20.1815 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.1263 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-5-25-12

Perfect score: 6774

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGLDPV 1249

Scoring table: BLOSUM62

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6161	91.0	1259	6	O18735 canis fam1
2	3141	46.4	1209	11	O9QX70 ractus norv
3	3112	45.9	1210	11	O9EP98 mus musculu
4	2730	40.3	1165	13	O9YH40 xiphophorus
5	2729.5	40.3	1137	13	O9W6F6 gallus galli
6	2292	33.8	1328	13	P79754 fugu rubrip
7	2000.5	29.5	1433	5	O9BIR9 anopheles g
8	1740.5	25.7	419	4	O9UK79 homo sapien
9	1739	25.7	367	11	O8R2X1 mus musculu
10	1720	25.4	729	15	O86712 avian rous-
11	1718	25.4	567	15	O86714 avian rous-
12	1697.5	25.1	412	4	O8WYV0 homo sapien
13	1653.5	24.4	962	15	O64895 avian eryth
14	1645	24.4	545	15	O85468 avian eryth
15	1503.5	22.2	655	11	O9WVF5 mus musculu
16	1487.5	22.0	643	11	O9ERV6 mus musculu

17	1282	18.9	1193	5	O9YIX8 ephydaria f
18	1188.5	17.5	1368	5	O23821 caenothabdi
19	1182	17.4	1717	5	O26566 schistosoma
20	1117	16.5	527	13	O90836 gallus galli
21	998.5	14.7	478	11	O9ES80 ractus norv
22	933.5	13.8	599	13	O9PBH2 gallus galli
23	906	13.4	165	4	O14256 homo sapien
24	887	13.1	176	11	O923V5 ractus norv
25	806.5	11.9	346	13	P11776 xiphophorus
26	778	11.5	435	5	O8SZW1 drosophila
27	754.5	11.1	311	13	O9PVE2 xiphophorus
28	752	11.1	1362	13	O9PVZ4 xenopus lae
29	738	10.9	1671	5	O9NVV5 biophthalari
30	723	10.7	149	6	O9BG66 oryctolagus
31	716.5	10.6	331	4	O9BUD7 homo sapien
32	705	10.4	1418	13	O93457 scophthalmu
33	696.5	10.3	1368	13	O8UW85 paratichthy
34	675.5	10.0	1369	13	O8UW86 paratichthy
35	670	9.9	1358	13	O73798 xenopus lae
36	658	9.7	1412	13	O8UW84 paratichthy
37	657	9.7	1472	5	O9U5A8 bombyx mori
38	651.5	9.6	1418	13	O8UW83 paratichthy
39	645	9.5	1245	13	O9YGH8 scophthalmu
40	643	9.5	1371	11	O9OCW4 drosophila
41	617	9.1	2144	5	O9VD34 ractus sp.
42	595	8.8	987	11	O91YMO mus musculu
43	593	8.8	935	4	O96LJ5 homo sapien
44	592	8.7	987	11	O99MR2 mus musculu
45	587.5	8.7	1036	4	O07912 homo sapien

ALIGNMENTS

RESULT 1	
O18735	PRELIMINARY; PRT; 1259 AA.
AC O18735	O1-CAN-1998 (TREMBlrel. 05, Created)
DT O1-CAN-1998	(TREMBlrel. 05, Last sequence update)
DT O1-UN-2002	(TREMBlrel. 21, Last annotation update)
DE ExB-2.	
OS Canis familiaris (Dog).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
OK NCBI_taxid=9615;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Yokota H.;	
RT "cDNA cloning of exB-2 from canine mammary gland."	
RU Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AB008451; BAA23127.1; -	
DR HSSP; P11562; IFGK.	
DR InterPro; IPR002048; EF-hand.	
DR InterPro; IPR000494; EGFR_L domain.	
DR InterPro; IPR000719; Euk_DKinase.	
DR InterPro; IPR002174; Furin-like.	
DR InterPro; IPR001245; Tyr_Kinase.	
DR InterPro; IPR004019; YLP_motif.	
DR Pfam; PF00757; Furin-like; 1.	
DR Pfam; PF01030; Recep_L_domain; 2.	
DR Pfam; PF02757; YLP_2.	
DR ProDom; PD006001; Euk_DKinase; 1.	
DR SMART; SMO0251; FU; 3.	
DR SMART; SMO0219; TYKc; 1.	
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.	
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW ATP-binding; Transferase; Tyrosine-protein kinase.	
SEQ SEQUENCE 1259 AA; 137969 MW; E3364D45C4ACD46 CRC64;	

Query Match 91.0%; Score 6161; DB 6; Length 1259;
 Best Local Similarity 90.4%; Pred. No. 0;
 Matches 1139; Conservative 43; Mismatches 66; Indels 12; Gaps 3;

QY 1 MELALCRWGLLIALLPGCAASTOVCT-----QYIKANSKFIGITELLVYGCQVVGNTL 54
 DB 1 MELAAWCRWGLLIALLPGCAAGTGVCTGDMKRLPASPETHLMDLRLHYGCGVVGNTL 60

QY 55 ELTYLPTNALSFLQDIOEVGVYLIANOVQVQLRIVRQQLFEDNYALAVLDNG 114
 DB 61 ELTYLPANASLSFLQDIOEVGVYLIANSQVQLRIVRQQLFEDNYALAVLDNG 120

QY 115 DPLNNTPTVPGASPGGLRELQRLSTELLKGGVLIQRNPOLCYOPTILMKDIFKNNOLA 174
 DB 121 DPLEGGIPAPGAAQGGRLRELQRLSTELLKGGVLIQRNPOLCHQDTILMKDIFKNNOLA 180

QY 175 LTLIDNRSRACHPQSPKCKSGRCWSESDCQSLTRVCGAGGACRGKPLPTCCHEQC 234
 DB 181 LTLIDNRSACRPSGACKHCKWAGSSGDCQSLTRVCGAGGACRGKPLPTCCHEQC 240

QY 235 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYTTDFEAMPNREGRYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYTTDFEAMPNREGRYTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLYCPHINQVTAEDCTQRCCKSKPCARVCYGLGMEHLREVAVTTSAN 354
 DB 301 YNYLSTDVGSCTLYCPHINQVTAEDCTQRCCKSKPCARVCYGLGMEHLREVAVTTSAN 360

QY 355 IOEFAGCKKIFGSLAFLPESFDPDASNTAPLOPEOLVFTLEITGYLYISAMPSTLP 414
 DB 361 IOEFAGCKKIFGSLAFLPESFDPDASNTAPLOPEOLRVFALBEITGYLYISAMPSTLP 420

QY 415 DISVFQNTLOVITGRLLHNGANSLTLOGISWLGRLSRELGSGLAIHNNHLCFVHTV 474
 DB 421 NLSVFQNTLRVIRGRVLDHNGAVSLTLOGISWLGRLSRELGSGLAIHNNHLCFVHTV 480

QY 475 PMDQLFRNPQALMTANRPEDECVGEGLACHQICARHCWGPRTQVNCSTQLRQEC 534
 DB 481 PMDQLFRNPQALMTANRPEDECVGEGLACHQICARHCWGPRTQVNCSTQLRQEC 539

QY 535 VEECRVLOGLPREYVNAHCLPCHBECQPNQSVTCFQPEADQCVACAHYKDPFCVARC 594
 DB 540 VEECRVLOGLPREYVNAHCLPCHBECQPNQSVTCFQPEADQCVACAHYKDPFCVARC 599

QY 595 PSGVAPDLSTYPMKFEPEDEGACQCPNCHSCVDLDDKCPAEQASLTSTVSNAVGS 654
 DB 600 PSGVAPDLSTYPMKFEPEDEGACQCPNCHSCVDLDDKCPAEQASLTSTVSNAVGS 659

QY 655 ILLVVVLGVFGLIKRRQOKIRKYTMKRLIQETELVEPLTPSGAMPNQAMRLKETEL 714
 DB 660 ILLVVVLGVFGLIKRRQOKIRKYTMKRLIQETELVEPLTPSGAMPNQAMRLKETEL 719

QY 715 RKVKYLGGAGFYVKKGIWIPDGENVKIPVAIKYLRNTSPKANKELLDAAVYAGVGSF 774
 DB 720 RKVKYLGGAGFYVKKGIWIPDGENVKIPVAIKYLRNTSPKANKELLDAAVYAGVGSF 779

QY 775 YVSRLLGICLTSTVQVLTQMLPYCCLLDHVENGRGLSGODLLNMCQIAKMSYLEDVR 834
 DB 780 YVSRLLGICLTSTVQVLTQMLPYCCLLDHVENGRGLSGODLLNMCQIAKMSYLEDVR 839

QY 835 LVHRDLAARNVLVSPNHVKITTDGLARLDDITETVHADGSKVPIKMALESILRRRT 894
 DB 840 LVHRDLAARNVLVSPNHVKITTDGLARLDDITETVHADGSKVPIKMALESILRRRT 899

QY 895 HQSDVMSYGVTVWELMTFGAKPYGIPARELPDLLEGERLPPPICTTIDVYIMYKCM 954
 DB 900 HQSDVMSYGVTVWELMTFGAKPYGIPARELPDLLEGERLPPPICTTIDVYIMYKCM 959

QY 955 IDSECRPFRFELVEFSGMARDPORFVIONEDLGPASPLDTFFYSLLDEDDMGDLVDA 1014
 DB 960 IDSECRPFRFELVEFSGMARDPORFVIONEDLGPASPLDTFFYSLLDEDDMGDLVDA 1019

QY 1015 EEYLVPQGGFFCQDPAPGAGVMYHRRSSSTRSGGDDLTLGLBPEEERAPSPPLASEG 1074

DB 1020 EEYLVPQGGFFCQDPAPGAGVMYHRRSSSTRSGGDDLTLGLBPEEERAPSPPLASEG 10

QY 1075 AGSDVFDGDLGMAAKGQSLPTHDPSPLQRYSEDPYPLPSETDGYVAPLTCSPQPEYV 11
 DB 1080 AGSDVFDGDLGMAAKGQSLPSPQPSPLQRYSEDPYPLPSETDGYVAPLTCSPQPEYV 11

QY 1135 NQPDVFRPQPSRPEEPPLAARPACTLER-----AKTILSPKRGVYQVAFAGAVENPE 11
 DB 1140 NQPEVWPQPALPEEPPLPSPRPACTLERPTLSPTKILSPKRGVYQVAFAGAVENPE 11

QY 1190 YLTPQGAAPCPHPPEAPSPAFNDLNYWDQPPRGAAPSTFFKGPPTAENPEYGLDVPV 12
 DB 1200 YLAPRGRAPCPHPPEAPSPAFNDLNYWDQPPRGAAPSTFFKGPPTAENPEYGLDVPV 12

RESULT 2
 ID Q9QX70 PRELIMINARY; PRT; 1209 AA.
 AC Q9QX70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN (1)
 RC SEQUENCE FROM N.A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RX MEDLINE=90258888; PubMed=2342465;
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Barp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Barp H.S.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN EMBL: M37294; AAF14008.1; -;
 DR HSEF; R11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00063; pkinase; 1.
 DR PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYFGK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR KAT-binding; Receptor; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEETFFCCIB7773 CRC64;

Query Match 46.4%; Score 3141; DB 11; Length 1209;
 Best Local Similarity 50.1%; Pred. No. 1,7e-228;
 Matches 636; Conservative 168; Mismatches 360; Indels 106; Gaps 25

QY 3 LAICRWGLLIALLPGCAASTOVCTGYIKANSKFIGITELLVYGCQVVGNTLTYLPTN 62


```

Db      15 LAALCAAGALEKKVCGGTSNR.LTOLGTEEDHFLSL-QRMFNNCVVLGNLEITYVOEN 73
Qy      63 ASLSFLADIOEVGVVLIANQVROVPLORLRIRYRGTOLEFDENVALADNGPLNNTTP 122
Db      74 YDLISFLKTIQEVAGVLIANTVRIPLNLTQIRGALIENTYALVALVSN----- 124
Qy      123 VTGASPGSLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDI FHKNNQALATLIDTNR 182
Db      125 -YGTNKTGLRELPRNRNIOETILIGAVRFSNNPILCNMETIOMRDI-VQDFELSNMSDVOR 182
Qy      183 S-RACHPCSPWCKSRKCMGSESEDCQSLTRTVACAGGA-RCKGPLRPDCCHECAGCGTG 240
Db      183 HLTGCPKCDPSCPRGSCMRGEENQCLTKIKICAQOQSRRCRSPBDCCNCCAGCGTG 242
Qy      241 PKHSDCLACLHFNHSGICELHCPALVTYNTDPESMNPGRXYTFGASCYACPYAVYLTST 300
Db      243 PRESDCLVCHFRDADCTKDTCTPPLMLYNPTVQMDVNPBGKXSFQATCYKCKPRRYVVT 302
Qy      301 DVGSCTLVCPRLHNOEYVAEDGTQCEKSKPCARVCGLCKEHLREVAVTSANIOEPAG 360
Db      303 DHGSCVRAAGPDYIEV-EEDGVSKCKKCDPCRCVKVCGIGIGEFKDTLSINATINIGHFYX 361
Qy      361 CKKIFGSLAFPLESPFDDDPASNTAPLQPEQLOVEETLEETGYLYISAMPDLSLPI:SVFQ 420
Db      362 CTAISGDHLILPVAFKXDSFTRTPLDPRLEILUKYTEITGTFLLIQAMPENMTDLHAE 421
Qy      421 NLQVIRGRILHNGVSLTLQGLGISWGLRSLRELSGSLAIHHNTHLCFVATVPDQLF 480
Db      422 NLEIRGRKHQGFSLAVVGLNITSLGLRSLKEISGDVILISGNMLCYANTINMKLIF 481
Qy      481 RNPQALLHTANREDECGEGLACHQLCARHGWGPGPCNCGQFLRGQCEVECV 540
Db      482 GTPNQTKIMNRKREKCKXATNHVCNPLCSGEGCBPPLDVCSCQNVSRGRGCDKCKI 541
Qy      541 LQGLPREVYNAHNLCPHPECQPNQSVTCFGEADQVACAHYKDPFCVARCSPGVNP 600
Db      542 LEGEPREFEVSECIQHPECLPQTMNITCTGRGPNCIKCAHYVGHPCVMTCPBGING 601
Qy      601 DLSYMPIKXEPDEGACQPCPINCITHSQVLDKCGPAERASRP-LTSIVSAVGLLVV 659
Db      602 ENNTL-VMKRADANNVCHLCHANCCTGCAGPLKGC--QOPEGPXKPSIATGIQGLLFI 658
Qy      660 VLGAVPGI-LIKRQOKIRKYRWRRLLOETELVEPLTPSGAMPNQAMILKETEIRKX 718
Db      659 VV-VALIGLFMRRRQLVRKTRIRLLQERELVEPLTPSGEARNQHLILKETEIRKX 717
Qy      719 VLSGAGFTYKGIWIPDGENVKI PVAIKYLRNTPSPKANKELIDRAYVMAGVSGPYSR 778
Db      718 VLSGAGFTYKGIWIPDGENVKI PVAIKYLRNTPSPKANKELIDRAYVMAGVSGPYSR 777
Qy      779 LIGICLTSTVOLTLQMPYGLLDHYENRGRGSGODLNMCMQIAKGMSTYLEDVRLYHR 838
Db      778 LIGICLTSTVOLTLQMPYGLLDHYENRGRGSGODLNMCMQIAKGMSTYLEDVRLYHR 837
Qy      839 DLAAARNVLKSPNHVKITDFGLARLDIDETEHADGKVPKIMWALBESILRRFTHSD 898
Db      838 DLAAARNVLKTPOHVITDFGLAKLIGAEKEHYHAGKVPKIMWALBESILHRYTHSD 897
Qy      899 VMSYGVTVMEIMFEGAKPYDGIAREIPDLLEKGEELPPRCITIDVYIMKCMNISE 958
Db      898 VMSYGVTVMEIMFEGAKPYDGIAREIPDLLEKGEELPPRCITIDVYIMKCMNIMAD 957
Qy      959 CRPRFRELVSFARMADPQRFVVIQ-NEDLGFASPLDSTFRSLLEDMDMDVLDAEEY 1017
Db      958 SRPKFRELILFESKMARDPQRYVLIQDGRMLPSTDSNFRALMEBEDMDVDVDAEY 1017
Qy      1018 LVPQOGFPCDPAPAGCMVHHNRHSSSTRSGGDLTLGLBSESEAPRSPPLASEGSGS 1077
Db      1018 LVPQOGF-----NPSPT-----SRPILSLSTANS 1043
Qy      1078 DVEDGDLGMGAAGKGLSLPTHPSPLOQYSEDPTVLPSET--DGVAVPLTCSPPPEVYN 1135
Db      1044 N-----SSTVACINRNSCWKEADAFLOQRSSPTSVTLTBDNIDTFL-----PYEYIN 1093

```

```

Qy      1136 QPDYRPPQSPREGRPLPARPACATLERAKTSPGKGVKQVDFAFGAVENPEYL-TRQ 113
Db      1094 Q-SVPKPPAGSVQVPHNQLPMP-----APGDHLHYQN--PSNVAVSNEYINTAQ 114
Qy      1195 GGAAPQHPPEPAPADNLYYDQ-----DP-----PERGAPSTFKGTPTAE 123
Db      1143 -----PTCLSSGFSALMWIQKSHQMSLDNDPDYQODFFPKAKPGRFKG-PTAE 119
Qy      1239 NPEYLGDDVP 1248
Db      1193 NAEYLRAVP 1202

```

```

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
ID Q9EP98;
AC Q9EP98;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
CN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egrf transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egrf transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_kinase; 2.
DR PRINTS; PR0109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00180; CYTOCHROME_C_UNKOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```

KW ATP-binding; Receptor; Transferase.
 SEQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;
 Query Match 45.9%; Score 1112; DB 11; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 2.7e-226;
 Matches 630; Conservative 169; Mismatches 356; Indels 118; Gaps 25;

11 LLLALLPGMA--STQVC-----TOYKANSKFIGITELLVQGCQVQGNLELTLP 61
 14 LITLCAAGALEEKVCQGSNRLTQGTGFEDHFLSL-QRMYNCCVVLGNLEITVQR 72
 62 NASLSFLQDIQGVQVLIANQVRQVQLQRLIVRTQLFEDNYALAVLDNGDPLNNT 121
 73 NYDISFLKTIQEVAGYVLIALTVERLENIQIIRGALYENTYALAIISN----- 124
 122 PVTGASPGGLRELQRLSTELIKGVLILQRPQLCYOPTILMKDI----PHKNNQALATL 177
 125 --YGNRTGLRELPRNQLQELLIGAVRSNNPILCNMDITQMRIVQVWMSMSMDL-- 180
 178 IDNRSRACHPSPCKSKRCWSESSDCCSLNRTVCAAGCA-RCKGPLPTDCCHEGCA 236
 181 --QSHSPSCPKCDPSCPGSCWGGEGENCQKLTXTIICAQCQSHRCGRSPSDCHNQCAA 238
 237 GCTGPKHSDCLACHFNHSGICEHLCPALVTYNTDTFSPMNPBGRYTFGASCVTACPN 286
 239 GCTGPRESDCLVCCQFOEATCKDTCPLMLYNLTTOQMDVNPBGKYSFGATCVKCKCPRN 298
 297 YLSTDVGSCTTLCPLHNGEVTAEQGTORCEKSKPCARVCYGLMEHLREAVATSANTQ 356
 299 YVVDHSGCVACGPDYEV--EEDGIRKCKKCDGPCRCKVNGIGIGFCKDTLSINATNIK 357
 357 EFAGCKKIFGSLAFIPESFDDGPASNTAPLOPELOVFEILEITGVLYSAPSDSLPDL 416
 358 HFKYCTAISDLHLIPVAFKDSSTRPPLDPRLELTLEKTVKELTGLLIQAMPDMMTDL 417
 417 SVFQNLQVIRRIILHNGAYSLTLQGLISWLGSRSLSELSGALIHNTLHLCFVATVP 476
 418 HAFENLEIIRGRTKHQGFSLAVGLNITSLGLRSLKEISDGVITISGNNNLTANTIMW 477
 477 DQLFNPHQALLHNPANREDECEVCEGLACHQLCAKRGCMRGPFOCCYNGCQPLRGECEV 536
 478 KKLFGTPNQCKKIMNNRAEKCKKAVNHVCMPLCSSESCWCPREPRDCVCSNVSRGCEV 537
 537 ECRVLDGLPREYVNAARHCLPCHPECPONGSVTFSGEADQCVACAYKOPREPCVARGCS 596
 538 KCNILEGEPRFVENSECICQHPCLPQANMITGTGPNQNCQCAHYIDGPHCVTCA 597
 597 GVKPDLSTMPITWKEPRDEGACQPCPINCTHSCVDLDDKCGCPAQRASPLTSTVSAVGL 656
 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGIQGCEVWPSGPKIPSIATGIVGL 656
 657 LVVVLGVVFGI-LTKRRQOKIRKXTMRLLQETELVEPLTPSGAMQNOQMRILKTELR 715
 657 LPIIV-VALGIGLPMRRRHIVRKETLRLLOERLVEPLTPSEAPRQAHRLKTEKER 715
 716 KKVVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYYWAGVSPY 775
 716 KIKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYYWAGVSPY 775
 776 VSRLLIGICTSTVOLVQLMPYGLLDHVENRGLRGSODLLWCMQIKMSYLEDVYL 835
 776 VCRLLIGICTSTVOLVQLMPYGLLDHVENRGLRGSODLLWCMQIKMSYLEDVYL 835
 836 VHRDLAARNVLVSKPNHVKITDFFGLARLIDITEYHADQGVPIKMWALESLIRRFTH 895
 836 VHRDLAARNVLVSKPNHVKITDFFGLARLIDITEYHADQGVPIKMWALESLIRRFTH 895
 896 OSDVMSYGVVWELMTGAKPCYQIGIPARELPDLLEGERLPQPPCTIVYIMYKCMWI 955
 896 OSDVMSYGVVWELMTGAKPCYQIGIPARELPDLLEGERLPQPPCTIVYIMYKCMWI 955
 956 DSECRPRFRELVSFSSMARDPQGFVVIQ-NEDLGPASPLDSTFFYSLDEDDMDGLVDA 1014
 956 DSECRPRFRELVSFSSMARDPQGFVVIQ-NEDLGPASPLDSTFFYSLDEDDMDGLVDA 1014

DB DADSRPKRELIILEFSKVARDPQRYVLVIOQDERHMLPSTFDSNPFYALMDEEDMEDVDA 101
 QY ERYLVPOQGFCTPPAPAGAGGMVHHRSSSTSGGSDLTLGLPSEBEAPSPPLAPSG 107
 DB DEYLVPOQGF-----NSPST-----SRTPLLSLS 104
 QY AGSDVFDDDLGMAKAGLQSLPTHDPSFLQRYSEDPVPLPSET--DGYVAPLTCSPQPE 113
 DB ATSN-----NSTVALINNGSGRVKEDAFLORYSDPFGAVTEIDNIDAF-----PVPE 105
 DB YVNC-SYKPRPAGSVQNPVYHNCPLP-----AGRDLYQN--PHSNAVGEYELN 114
 QY TPQGAAPQHPHPAPAPAPNLYYMQ-----DP-----PERGAPSTFKGT 123
 DB TAO-----PCLSSGFNSPALMLQKSHQSLNDPQODFFPKETKPNGIFKPG- 119
 QY 1236 TAENPEYLGIDVP 1248
 DB 1191 TAENAEYLRAVP 1203

RESULT 4
 Q3YH40 PRELIMINARY; PRT; 1165 AA.
 AC Q3YH40;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 NX NCBI_TaxID=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RA Altschmed U., Schachtl M.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.";
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Schachtl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53471; AAD10500.2; .
 DR HSSP: P13362; 1FGK.
 DR InterPro: IPR000345; CytC heme bind.
 DR InterPro: IPR000494; EGFR L domain.
 DR InterPro: IPR000719; Euk_PKinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyr_PKinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR Prodom: PD000001; Euk_PKinase; 1.
 DR SMART: SM00261; Fv; 3.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00343; GRAM_POS ANCHORING; UNKNOWN 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM; 2.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

SEQ	SEQUENCE	1165 AA; 129614 MW; 767EE38D8771A74E CRC64;
Query Match	40.3%; Score 2730; DB 13; Length 1165;	
Best Local Similarity	45.8%; Pred. No. 2e-197;	
Matches	582; Conservative 163; Mismatches 383; Indels 142; Gaps 30;	
QY	1 MELALCRWGLLALDPG-AAST---QVC-----TOYIKANSKFTIGITELLVGGCGVQV 51	
DB	4 LELLEL-----LLLLLLLSIGRCSTDPDRKVCQGTSMQMTMLDHHYLMKMK-MYSGCNVVL 58	
QY	52 GNLLETPLTNASLSFLODIQEVQGYVLIHNOVROVPLORLRIYVGTQGLFEDNVALATL 111	
DB	59 ENLEITTTQENODLSFQSIQEVGYVLIHMANVESTIPLVNLELIRGQVLEIGNFTLLWM 118	
QY	112 DNGDPLNNTTPVTGASDGLRELQSLTEILKGYVLIQRNPOLCYQDTILMKDIFHKAN 171	
DB	119 SNYCK-NPSSP--DVYQVGLKQQLSNLFTIELSGVKVSHNPLCNVETIIMMDIVDKTS 175	
QY	172 QLATLIDTNRSRACHPCSPMGKSGRCWGSSEDCQSLTRTVAGGC-ARCKGPLPTDCC 230	
DB	176 NPTMNLIPHAPEHQCCQCPGCVNGSCWAPGPHCQKFTKLCAEQCNRRCRGPRIDCC 235	
QY	231 HEQCAAGCTGPKHSDCLACLFHNSGICEILHCPALVTYNTDTFESMPNDEGRYTFGACV 290	
DB	236 NEHCAGGCTGPRAITDCLACRDFNDGCTCPRKIYDISHQVVDNINIKYTGACV 295	
QY	291 TACPYNLTSDVQSCITLVCPLHNOEYTAEDGTORCEKSKPCARVCYGLMEHLREAV 350	
DB	296 KECPSNYVTE-GACVASCAGMLEVD-ENGKSKCPKCGVCPKVCDDGIGISLNTIAV 353	
QY	351 TSANIOFAGCKKIFGSLAFLPESPDGDPASNTAPLOPQLOFTEILEITLYLISAMP 410	
DB	354 NSTTIGSFNCTKINGDIILNRSFEDDPYKIGPDPHMLNLTIVKSTLYLIMMP 413	
QY	411 DSLPDLISFQNLQVIRGRLHNGAYS-LTLQSIGISMLRSLRELGSGLAIHNTILC 469	
DB	414 ENMTLSLVSFQNLRIIRGRTFSRGSFVVYQVSHLQWLGLSLKEVSAQNVILKTPQLR 473	
QY	470 FVHTVPMDOLEFRPHQALLHTANRPEDECYGBELAHQICAGHCHMGPRPTCCVNCQFL 529	
DB	474 YASTINMRRLFRSEDSQIEYDAT-----EQCTNNESDGCWGPPTCVASCLHVD 526	
QY	530 RQGEVCEECRVLQGLPREYVNAHCLPCHPECCPONGSVTFCGPEADQVCAAHYKDPF 589	
DB	527 RGRGVASCNLLQGEPRBAQVDGRVCVQHQECLYQDSLTICGPRGANSKAAHQDQBP 586	
QY	590 CVARCPGKVPDLSTYPIWKFPEBEACOPCPINCHSCVLDLDDKGCPEAGCASPLTSIV 649	
DB	587 CIRCPHGMIGDDTL-IMKYADKMGQCCPCHQNCQSGSPGLSGCRGD-IVSHSLAV 644	
QY	650 SAVVGLLVVLVGVFGLIKRQOKIRKYTRRLIQLTELEVPLTPSGAMPNQMRIL 709	
DB	645 GLVSGLLITYVALLVILRRRIR-RKRTIRLLQELVEPLTPSQADNQAFIL 703	
QY	710 KETELRKVVLGSGAGFYVYKGIWPDGNVXIPIVAKYLRENTSPKANKELDEAYVMA 769	
DB	704 KETEFKXDRVLGSGAGFYVYKGIWPDGNIRIPIVAKYLRATSPKNOEVLDEAYVMA 763	
QY	770 GVGSPYVSRLLGLCLSTVQVLTQMLPYGCLLDHYEENRGRLGSDLLMCMQIAKGSY 829	
DB	764 SYDHPVCRLLGLCLSAVQVLTQMLPYGCLLDYRQOERICGQMLNMCQIAKGNV 823	
QY	830 LEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGVPIKMALESIL 889	
DB	824 LEEERHLVHRDLAARNVILKPNHVKITDFGLSKLLTADKEVQAGKGVPIKMALESIL 883	
QY	890 RRRFTHQSDVWSYGVVWELMTGAKPYGIPARELPDLLEGEELPPPTCTIDVYIM 949	
DB	884 QMTYHQSDVWSYGVVWELMTGSKPYGIPAKELASYLEGEELPPPTCTIEVYMI 943	
QY	950 VKCMIDISECRPFRELVEFSRMAADPQRFVYIYQNEIDLGASPLDSTFYSLLEDDMG 1009	
DB	944 LKCMIDIPSSRPFRELVEFSQMAADPFRVYIYG---NLPSPEDRLFLRLLSSD-- 998	

QY	1010 DLYDAEEYLVQQCFPCPAPAGGVHHRSSSTRSGGDLTLGLEPSEEARSP 106	
DB	999 DLYDADLYL-----RYKRIN-RQS-----EPC 102	
QY	1070 APSEGASDVFDGDLGMAKGLQSLPTHDPSPFLQRYSDPTV-PLPSETDGYVAPLTS 112	
DB	1022 IPRNGH-----PVRENSIALRYISDPTQNALEKLDGH----- 105	
QY	1129 PQPEYVNPDRPQ-----SPRE-----GPLP-AARPAATLERAKTISPGRNV 117	
DB	1055 ---EYVQPSSETSSRLSDIYNNYEDLTDGWPVLSQGEAFETNFSREYVNTQNSI- 111	
QY	1176 KDVFAGGAVNEBEYLTLPQGAAPRPPAPFAPFADNLYWDDPPEGAPPSTFGTP 123	
DB	1111 --PLVSGSSMDPDDY---QAG-----YQAA------LPQTGALTGNMFLP 114	
QY	1236 TAENPEYGL 1245	
DB	1147 AANLELYGL 1156	

RESULT 5

ID Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6; 01-NOV-1999 (TEMBLrel. 12, Created)

DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Receptor tyrosine kinase (Fragment).

OS ERBB4.

OC Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

NCBI_TaxID=9031.

RN [1]

RE SEQUENCE FROM N.A.

RC TISSUE=HINDRAIN;

RX MEDLINE=9263203; PubMed=10328884;

RA Dixon M., Lumsden A.;

RT "Distribution of neurogulin-1 (nrg1) and erbB4 transcripts in

RL Mol. Cell. Neurosci. 13:237-258(1999).

DR EMBL; AF121963; AAD31764.1; .

DR HSSP; P11362; IFGK.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_dkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001368; TNFR_c6.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; kinase; 1.

DR Pfam; PF01030; Recep_L_domain; 1.

DR Pfam; PF02757; YLP_2.

DR PRINTS; PR00109; TYRKINASE.

DR ProDom; PD000001; Euk_dkinase; 1.

DR SMART; SM00261; FU; 3.

DR PROSITE; PS00219; TYRK; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.

DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.

DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.

KM kinase; Tyrosine-protein kinase.

FT NON_TER

SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 40.3%; Score 2729.5; DB 13; Length 1137;
 Best Local Similarity 47.3%; Pred. No. 2.1e-197;
 Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

155 LCYQDTIIMKDIHFHKNQALTLIDTNRSRACHPCSPKSGRCWGSSEDCQSLTRTVAC 214

Db	3	LCFADTIIHMODIVRNPMASNFTLVPNTGSSGCGHCKSCIG--RWMGTENHNCQTLITTVIC	61
Qy	215	AGGC-ARCKGRLPTDCCHEQCAAGTGKHSDCIACLFHNSGICELHCPALVTYNTDTF	273
Db	62	AEQCDRCYGPVSDCHRECAAGSGGSKNDCAFAMNFDSGACVQCQOTFEVYNTTF	121
Qy	274	ESMPNREGRTFGASCYACFPYNYLSTVSGCTVCPHLNDEVAEDGTORCEKSGPCA	333
Db	122	QLEHNHNAKYTYGACVCKCHNFV-VSSSCVACACSSKMEV-EEGIGKCKRCPTDICP	179
Qy	334	RVCYGLGMEHLREVAVTSANIQEFAGCKTIFGSLAFIPESFDGDPASNTAPLQPEQLQY	393
Db	180	KACDGIQSGSLVSAQTVSSNIDKFNCTKINGNLIFLVGIHDDPHYTAIAINPEKNTI	239
Qy	394	FETLEITGYLYISAMPDLPDLVSFQWLYIRGRILHNGAVSLTQGLISWLGRLSR	453
Db	240	FQTVREITGYLQINQMFEMTDFRVSNTLVITGRALYSGLSLILKQGITSLQFQSLK	299
Qy	454	ELGGLALIHNTLGFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGLAGHQLGARGH	513
Db	300	QISAGNIYITDNSNLCYHTVNMWTSLEFSTPSQKTVIHRNKKAENCTADGMVCHLSSDG	359
Qy	514	CMGPGTQCVNCSQPLRQECVEGCVLQGLPRVYVNAHCLPGHPCCP-QNGSVTCG	572
Db	360	CMGPEPDCLSCKRIIRGRTCTESCNLDGFEFFAVGAVCMCEDPQCEKMEENMTTCYG	419
Qy	573	PEADQCVACAHYKDPFPCVACRPGVXPDLSYMEIMKFPDEGACDPCPINTHSCVDLD	632
Db	420	FGPHCTKCFHFKDQPNCEKCPDGLQANSF--IFRYADEBDEHCHCHNCQCGRGR	477
Qy	633	DKGC-----PAEQRASPLTSVSAVY-GILLVVVLGVVFGILLRROOKTRKXT	680
Db	478	SHDCIYFWTRQSTLPHAR--TPV--TAAAGTIGLFTIIVMGTLFAYVARRKSIK-KKRA	533
Qy	681	MRLLQETELVEPLTPSGAMPNQAOMRILKETELRKYKVGSGAFVYVYGIPIPGENV	740
Db	534	LRRL-ETELVEPLTPSGTANQAOLILKETELRKYKVGSGAFVYVYGIPIPGENV	592
Qy	741	KIPVAIKVLENTSPKANKELLDAAVYVAGVSGVYVRLIGICTSTVQLVTLQMPYGL	800
Db	593	KIPVAIKILNTTPKNAVEMFMDALIMASMDHPLVRLGVCISPTQIVTLQMPHGC	652
Qy	801	LDHYRENRGLSGODLNMCMQIAKMSYLEVRLVLRDLAANVLYKSNHNTKIDEG	860
Db	653	LDYVHEKDNIGSQLNMCVQIAKMMYLEBRLLVLRDLAANVLYKSNHNTKIDEG	712
Qy	861	ARLLIDETEVHADGKVPKIMWALSILRRRFTHSDVMSYGVTVWELMTFGAKPYDGI	920
Db	713	ARLLEGEKEYNADGKMPKIMWALECIHYRKFTHQSDVMSYGVTVWELMTFGAKPYDGI	772
Qy	921	PARLIDLEKGEPLPOPICTIDVYIMIKMWIDSECPREFELVSESRMARPPOR	980
Db	773	PTRELIDLEKGEPLPOPICTIDVYIMIKMWIDSECPREFELVSESRMARPPOR	832
Qy	981	VVIONED-LGPASPLDSTFYRSLLEDDMDGLVDAEELVLPQGFCCPPAPAGAGMWH	1039
Db	833	LVIQDDMKLPSBNDKFPQNLLEDEDLDDMMDAEELVLP-QAFNIPPIYTSKRIDS	891
Qy	1040	RHSSSTRSGGDLTLLEBSEEFAPS--PLAP-SEGAGSDVFDGDLGMAKGLQISLP	1096
Db	892	NRNOFYVRDGGYAAEGV-PMFYRAPGCIIPAPVAGATAEIFEDTCNGTLRKQVATL	950
Qy	1097	THDPSLQRYSEDTVPLPS-----ETDGYVAPLTCSPQPEYVQOPVRRQPPSPRG	1149
Db	951	AKEDSSTQRYSDPTVIFRERVIKGLDEDEGWTTPRDKRKIDYINLPVENFVSRKKG	1010
Qy	1150	PLPAA-RPAGATLERAKTLPKGNKVGVF-----AFGAVENPEYLPQCGAARQ	1200
Db	1011	DLQAVNDPEYHN-----APNGQPAEDEVYVNEPLYLNTFANTLENAEYL-----K	1055
Qy	1201	PHPPAFSPAFDNLVYWDQDPPERGA--PSTFKGTPT-----AENPEYL	1243
Db	1056	NNLPEKAKKAFDNDYNNHSLPPRSTLQHHDYLOEYSTKFKYKQNGRIRPIVAENPEYL	1114
Db	RESULT 6		
Qy	79754	PRELIMINARY; PRT; 1328 AA.	
Db	AC	P79754	
Qy	DT	01-MAY-1997 (TRENBLREL.03, Created)	
Db	DT	01-MAY-1997 (TRENBLREL.03, Last sequence update)	
Qy	DT	01-JUN-2002 (TRENBLREL.21, Last annotation update)	
Db	DE	Etdb3.	
Qy	OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).	
Db	OC	Fuguryctia; Metazoa; Chordata; Craniata; Euteleostomi;	
Qy	OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
Db	OC	Tetraodontiformes; Acanthopterygii; Perciformes; Tetraodontidae; Takifugu.	
Qy	OC	NCB1_Taxid=31033;	
Db	RN	(1)	
Qy	RP	SEQUENCE FROM N.A.	
Db	RX	MEDLINE=99177347; PubMed=10077531;	
Qy	RA	Geliner K., Brenner S.;	
Db	RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu	
Qy	RT	rubripes.";	
Db	RL	Genome Res. 9:251-258(1999).	
Qy	DR	EMBL; AF056116; AAC34391.1; -	
Db	DR	HSSP; P11362; IFGK.	
Qy	DR	InterPro; IPR000494; EGFR_L_domain.	
Db	DR	InterPro; IPR000719; Euk_Dkinase.	
Qy	DR	InterPro; IPR002174; Euk_Dkinase.	
Db	DR	InterPro; IPR001245; Tyr_Dkinase.	
Qy	DR	Pfam; PF00757; Furin-like; 1.	
Db	DR	Pfam; PF00069; Dkinase; 1.	
Qy	DR	Pfam; PF01030; Recep_L_domain; 2.	
Db	DR	Pfdom; PD000001; Euk_Dkinase; 1.	
Qy	DR	SMART; SMO0261; Fu_3.	
Db	DR	SMART; SMO0219; Tykic; 1.	
Qy	DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
Db	DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
Qy	KW	ATP-binding; Transferase.	
Db	SEQUENCE	1328 AA; 148613 MW; A33039258B647E9 CMC64;	
Qy	Query Match	33.8%; Score 2292; DB 13; Length 1328;	
Db	Best Local Similarity	41.28; Pred. No. 3.3e-164;	
Qy	Matches	510; Conservative 148; Mismatches 397; Indels 184; Gaps 29	
Qy	44	YQGQVQGNLELTYLPTNASLSFLQDICEVQGVLIHANQVQVPLRIVRGTQLE	103
Db	51	YKGEIIMGNLEITQIESNMPDFLKTIRVTVYVLIAMNHPEIFLQGLRVYKNSLVE	110
Qy	104	DNYLAVLDNGDPLNNTTPVTGASPGGLRELRSLUTEILKGGVLIQRNPOLCYODTILM	163
Db	111	RRFALSVFLN-----YPKDG--PSGLNQLGIMNTEILIDGGVQLINNKVLRVYGMVWY	161
Qy	164	KDIFKKNQALTLTIDTNRBACHPSPCKSGRSGESESSEDOSLTRIVCAQGC-ARCK	222
Db	162	RDLI-RKNDAPRIETQNGRNGVCH---KSC-GNYCWGPKGQQLITKTVCAPOCNDRCF	216
Qy	223	GPLPTDCHEQCAAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREG	282
Db	217	GTSFRDCHIECAAGCKGPLDTCFACRLFNDSGACVPCQPTLIYNKQTFQMEYNPAK	276
Qy	283	YTFGASCVTAQCPYNYLSTVSGSTLWCPHLNQEYTAEDGTOR-CEKGSFPCARVYCGLM	341
Db	277	YQGISIVSQCPYHAF-VDSGSCVSYCPDPKMEV--ERSQRCQCELCGSLCPKVEGTGA	333
Qy	342	EHLREVAVTSANIOEFAGCKTIFGSLAFIPESFDGDPASNTAPLQPEQLQYFETLEBIT	401
Db	334	E--QRTVDSNIDSFINCTKIQSLHLTVLGIIDDPKVPPLDAKTLVFRVREIT	390
Qy	402	GVLISAMPQSLDLSVFQWLYIRKRIHNGAVSLTQGLISWLGRLSGRLSGGLAL	461
Db	391	DIINTQSWPEKELNDLSVFSLLTIQGRSLFKRFSIMVMKIPTLTSLGRLSRLSISDGSVY	450

```

QY 462 IHANTHLCFVTVFMDDLFNPH-QALLHTANRPEDECVGBGLACHOLCARHCWGPPT 520
DB 451 ISONALCYHHTVMTOLFGRSRVRANSLNSRPMACVADGRCVPLCDSDSCWCFGPD 510
QY 521 QCVNCSQFLRGQCEVEGRVLOGLPREYVABH-CLCEHCEPCOPONSVCYCFGPADGCV 579
DB 511 QCLSCRNSRHCCTVAGCHPNSGIPREFAGANGVACVACHECKRQJGKASCTSPGADBCW 570
QY 580 ACAHYKDPFECVAPCPSGVKPDLSYMPIMKFPDEGACOPCPINCTHSQVLDLDDKCPAB 639
DB 571 ACTKFRDGPYCMSCSPAGVN-DGEKGLIFKFPNNEGCEPCHQCTGGCGPGLNDC--- 626
QY 640 QASPLTISVAVGILLVYLVGVF-----GILIKRQOKIRKXTMRLLQELIVE 692
DB 627 LEAARLTISSQGITGIALGVPAAGILFCLVFLFGLMLYHGLAIRRKAMRYLESESE 686
QY 693 PLTPSGAMPNOAMRILKETELRKVKYLGSGAFGTYYKGIWIPDGENVKIPVAIKYLRN 752
DB 687 PLGP-GEKGTVYHARILKPSDLRKIKPLGSGVFTGSKGFWIBGETVKIPVAIKTIQPS 745
QY 753 TSPKANKELDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGLDHYENRGRIG 812
DB 746 SGRQTFEITDHLISMGLDHPYIVRLGICPGTCLQVLVQLSHSLLEHTRQHTSLD 805
QY 813 SQDLNMCQIAKXMSYLEDVRLVHRLAARNVYKSPNHVKITDPLARLIDITEYH 872
DB 806 PQRLLNMCVQIAKMYLIEHRVYHKLARNLILKNDYQVQSLSDYGVADLIPDDKXYV 865
QY 873 ADGKVPKIMMALESILRRFTHQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKG 932
DB 866 YSEKTPKIMMALESILFRRYTHQSDVWSYGVYVWEMMSGAPRYASVQOEYPSVLEKG 925
QY 933 ERLPOPICITIDYVIMVWKCMVIDSECRPRELVSFESMAADPQRFVYIQWEDIGPAS 992
DB 926 ERLSOPALCTIDYVIMVWKCMVIDENRPTFKELASDFTMAADPPRYLYIRNEG----- 980
QY 993 PLDSTFYRSLLEDMDKDLVDAEYIVPQGFPCPPAPGAGGVHHRHSSTRSGGAD 1052
DB 981 -----EDSGMGEFL-----RRGSR---GL 997
QY 1053 LTLGLEPSEEFAPSPPLAPSGAGSDVFDGDLQMG---AAKGIQSLPTHDPSPLO----- 1104
DB 998 LEADLEDEDEE-----GLGDRFATPQLQPSPSWSTSPSQINSYM 1036
QY 1105 -----RYSEDPVPLPSETDGYVAPLTCSPQ- EYVNO-----PDVAPQPP 1144
DB 1037 VMTQLRYD-----FAVSGGHIGYLPMSDPYDTRQLMYQSRSLSVRTLPDRSAFR 1090
QY 1145 SPREGPL--PAAPAGATLEBAKTLSPGKNGVADVAFAGGAVENPEYLLPQGAAPDPH 1202
DB 1091 SSRRAELCEDGAGCAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1203 PPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPE 1241
DB 1123 -----QQRKLTASSPSFSFTWMADEDE 1146

```

```

RESULT 7
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anophelinae; Neophelinae; Anophelinae; Anophelinae;
CX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-SUA;
RA Lysate G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DB EMBL; AJ301655; CAC35008.1; -.
DB HSBP; F11362; 1FEK.
DB InterPro; IPR000345; Cyt_c_heme_bind.
DB InterPro; IPR000494; EGFR_L_domain.
DB InterPro; IPR000719; Euk_Pkinase.
DB InterPro; IPR002174; Furin-like.
DB InterPro; IPR002290; Ser_thr_pkinase.
DB InterPro; IPR001245; Tyr_pkinase.
DB Pfam; PF00757; Furin-like; 1.
DB Pfam; PF00069; Pkinase; 1.
DB Pfam; PF01030; Recep_L_domain; 2.
DB PRINTS; PRC0109; TYRKINASE.
DB ProDom; PD000001; Euk_Pkinase; 1.
DB SMART; SM00261; FU_7.
DB SMART; SM00220; S_TKC; 1.
DB SMART; SM00219; TYRK; 1.
DB PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DB PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DB PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DB PROSITE; PS00103; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D8896724F07 CRC64;

```

```

Query Match 29.5%; Score 2000.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No. 42e-142;
Matches 464; Conservative 191; Mismatches 383; Indels 385; Gaps 36
QY 44 YQCGVAVQNSLELTYPNTASLFLDIOGVQYVLIHANGVQVPLQRLIRYGTQLF- 102
DB 25 YTNCTYDGNLEITWIONITDNLNPLQHIREVGYVLSLYDLQVILPRQITRGHTTFK 84
QY 103 -----EDNYALAVLDNPGPLANTTTPVVGASPGRLQSLSTLILGCVLIGNPOLCVO 158
DB 85 LNKWEANYGLFV-----SSHNKTLLEPLRLDITLGSVGFNNVNLCHN 128
QY 159 DTLMKDI-FHKNNQALTLIDTRSRACHSPCKGSRCKGSESSDQSLTRTVACAG 217
DB 129 KSIWBEILAPQSMGYTFNFSPEVCPCHPSCSEVG-CWEGAHNQRPSKLNCSQ 187
QY 218 CA--RCKGPLPTDCCHQCAAGCTGPKHSDCLALPHNHSIGTELCPALVTYNTDFES 275
DB 188 CSQRCGPKPRRECHCFCAAGCTGPTOSDCLACKNFYDGVCKECPMPQIYNPTNYFW 247
QY 276 MPNPEGRTYGASGVTPACPYNYISTVGSCTVCPILNQEVTAEQTCRCKSKCAEV 335
DB 248 EPNDDGKAYGATCYRKP-EHLIKNGACVRCCKPKKPPQNSE-----CVPCKGVCPXT 301
QY 336 CYGLGMEHLREVAVTSANIQEFAGCKKIFGSLAPLPESFDDGPASNT-----APLQ 387
DB 302 CPGEGLVH-----SDNIGNYKDCITIEGSLDIDQSPDFQVYTFNFSQPRYIKID 353
QY 388 PEOLQVETLEETITGYLYISAMDSLDLSVFNQLOVINGRIILHNAV-SLTLOGIGIWM 446
DB 354 PDRLEVSTYKEITGFINTQAHHPNTTLYNFRNLELVGGRQKENLFASVYIVKTSLS 413
QY 447 LGLSLRELGSGLALIHNTHLCFVTVFMDDLFNPHQALHTANRPEDECVGBGLACH 506
DB 414 LEUKSLKRVNVSQIVILENSDLDFVEDIMSEIKSSDHEWVQKRNNTTECHHEMBES 473
QY 507 QLCARHGCMGPGTQCVNCSQFLRGQCEVBECEVLOGLPREY-VNARHGLPCHPBCOPON 565
DB 474 EQCSKACQWKGEGQCECNVYKXKCLDSCK--SLPRLYSVSKTGCDGHOEKC-D-- 528
QY 566 GSVTCFPGPADQCAVAAHYKDPFCVAPCP----- 595
DB 529 ---FCYGNEDNCGSCMNVYDGFCAVCEPPTKHAMNGTCINCHKTVCGRPRTIAPD 585

```

```

QY 596 -----SGVKPDLSTYMPIMKFPD-----EEG----- 615
Db 586 GCISCKAIIGSDAKIERCLMKDESCPGYSDVYLQEGGLKOLSGAVCRKCHPRCK 645
QY 616 -----ACQPCPINCT-----HSC 628
Db 646 CTGYGFHEQCECTGYKKGEQCEDECEQDFYANETRICLPCHQECGCHGLDDHEC 705
QY 629 VDL-----DKCPAEQ-----RASPLTSI 648
Db 706 RNLKLFEGDPYNATFTTCVSNCPASHYKFRPGAGIKGYCSADSQSLREPOTQV 765
QY 649 VSAVVGILLVVLGVVFGI---LIRROQKIRKTYRRLIETELVEPLTSGAMPNQAQ 705
Db 766 KIVMGVMAILLICVFGIAPVLFGRHKKKDAVMTALAGCDESEPLRSNVGPNLTJK 825
QY 706 NRILKETELRKVYVSGAGFTVYKGIPIPGENKIPVATIKULRENTSPKANEILLDEA 765
Db 826 LRIIKAEIRGCVLQMGAFGRVFGVMMPEGSVKIIVALKVLEMSGSESSKEFLEEA 885
QY 766 YVAGVSPYVSRLLIGICTSTVQVLTQMLPYGCLLDHVRNQRGLSODLLNMQOIAK 825
Db 886 YMASVEHENLLKTLAVCMVTSQMLITQMLPLGLDVRNNKDKIGSKALLMNSTQIAR 945
QY 826 GMSYLEDVRLVARDLAARNVLYKSPNHKIIDFGIARLIDIDETRYADGKVPYIKMAL 885
Db 946 GMAVYEBERLVARDLAARNVLYQTPSCVKITVFGIATKLPDSDEYAAAGKMPKIMLAL 1005
QY 886 ESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGIAPAEIPDLLEKEKERTPOPICITDV 945
Db 1006 ECIRHRRVFTSKSDVAFGITITWELTYGARVYENVPKADVPBELLIGHKLPQDPCGLDV 1065
QY 946 YVIMYKCMWIDSECPREFRELVSERAPDORFVVIQNEDELPAASPLDSTFYRSLLED 1005
Db 1066 YCILLSCWVLDADAPTEFKOLAEFTAEKARDBGRYLM-----PGDKFMELPSYTN 1116
QY 1006 DDMGDLV-----DAEYLVPOQGFPCPDPAAGCM 1036
Db 1117 QDEKDLIRTLAPVMAAAAAAAAAAGASVNDVPTIAETDELQKTPSITMLPEPSA--- 1173
QY 1037 VVHHRSSSTSGGDDLTLGLEPSEEEAPRS-----PLAP---SEGAGSDVFDGLGMG 1087
Db 1174 -----VEPS-DEMPKSLRYCKXPLKPDDETQHGKEV-----GVG 1207
QY 1088 AAKGQSLPTHDPSPLOQYSEDPVPLPSEITDGVAVLTCSPQREYVNGQDVAPQPSR 1147
Db 1208 GIR-----LNLPLDEDDYLMPTCOSQ---NQS----- 1231
QY 1148 EGPLPAPAPAGATLERAKTISPGKNGVVKVFAFGAVENPEYL-----TPQSGAA 1198
Db 1232 -----TPG-----YMDLIGVPAVDNPEVLMGSTQAIAGLAGGSMG 1267
QY 1199 PQPHPPAPSPAFDNLVYWDPPERGAPESTFGTATNPE 1241
Db 1268 --PHTP-----PPTVNGMPTHQHSQ 1287

```

RESULT 8

```

QY 09UK79 PRELIMINARY; PRT; 419 AA.
AC 09UK79:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxId=9606;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;

```

```

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RL autoinhibitor.";
RN Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
(2)
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB177761; AAD56009.2;
DR InterPro; IPR00494; EGR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like.
DR Pfam; PF01030; Recept_L_domain; 1.
DR SMART; SM00261; Fu; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLBE347E2D030C CRC64;

```

```

Query Match 25.7%; Score 1740.5; DB 4; Length 419;
Best Local Similarity 84.3%; Pred. No. 3,2e-123;
Matches 332; Conservative 9; Mismatches 40; Indels 13; Gaps

```

```

QY 1 METALCGWGLLALPPGASTQVCT-----QYKANSKFIGITELLYGCGVVGNL 54
Db 1 METALCGWGLLALPPGASTQVCTGTDMKRLPASPEHLMRLHYGCGVVGNL 60
QY 55 ELTYLPTNASLFLQDIOGVGYLIANQYQVPLQRLRIVRSTQLFEDNYALAVLDNG 114
Db 61 ELTYLPTNASLFLQDIOGVGYLIANQYQVPLQRLRIVRSTQLFEDNYALAVLDNG 120
QY 115 DPLNNTVTGASRGLELQSLTELKGVLIQNPOLCTQDTILMKDIFKXNOLA 174
Db 121 DPLNNTVTGASRGLELQSLTELKGVLIQNPOLCTQDTILMKDIFKXNOLA 180
QY 175 LTLIDNRSRACHPCSPMKCKSRGWCSSSEDCQSLRTVACGACGKGPLPDCHEQC 234
Db 181 LTLIDNRSRACHPCSPMKCKSRGWCSSSEDCQSLRTVACGACGKGPLPDCHEQC 240
QY 235 AAGCTGPGHSCCLACHPNHSIGIELHCPALVTYNTDFESMPNREGRYTFGASCYTACP 234
Db 241 AAGCTGPGHSCCLACHPNHSIGIELHCPALVTYNTDFESMPNREGRYTFGASCYTACP 300
QY 295 YNYLSTVGSCTLVCPHNOVTAEDGTQCEKSKRCARVCGLGHEHREYRAYTSAN 354
Db 301 YNYLSTVGSCTLVCPHNOVTAEDGTQCEKSKRCARVCGLGHEHREYRAYTSAN 355
QY 355 IQEPAGCKKIFGSLAPFESFDGDPASNTAPLP 388
Db 356 LRMQPG--PAHPVLSFLRPSGMDVSAFYSIPLAP 387

```

RESULT 9

```

QY 08R2X1 PRELIMINARY; PRT; 367 AA.
AC 08R2X1:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 111
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AA027080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

```

```

Query Match 25.7%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 3,4e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0.

```

```
QY 883 MALSSILRRRFTHSDVWSYGVYWEIMTGANPYDGIIPAREIPDLLEKGERLPQPICT 942
Db 1 MALESILRRRFTHSDVWSYGVYWEIMTGANPYDGIIPAREIPDLLEKGERLPQPICT 60
QY 943 IDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFVYIQNEDLGSPASPLDSTFYRSL 1002
Db 61 IDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFVYIQNEDLGSPASPLDSTFYRSL 120
QY 1003 LEDDDMGDLVDAEEYLVPOQGFPCPDPAFGAGMWHHRHSSSTRSGGDLTLGLEPSEE 1062
Db 121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAFGAGMWHHRHSSSTRSGGDLTLGLEPSEE 180
QY 1063 EAPRSPLAPSEAGSDVDFDDLVAGVTGSLQSLPHDLSPQRYSEDPTLPLPETDGYV 1122
Db 181 EAPRSPLAPSEAGSDVDFDDLVAGVTGSLQSLPHDLSPQRYSEDPTLPLPETDGYV 240
QY 1123 APLTCSQPPEYVNOPDVRPQPSFREGPLPAARPAATLERAKTLSPGKNGVQDVFAFG 1182
Db 241 APLTCSQPPEYVNOPDVRPQPSFREGPLPAARPAATLERAKTLSPGKNGVQDVFAFG 300
QY 1183 GAVENPEYLPQGGAAQPHPPAFSPAFDNLTYWQDPPERCAPSTFKGPTAENPEY 1242
Db 301 GAVENPEYLPQGGAAQPHPPAFSPAFDNLTYWQDPPERCAPSTFKGPTAENPEY 360
QY 1243 LGLDVPV 1249
Db 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; Pubmed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1.
DR Pfam; PF02813; Retro_M.1.
DR ProDom; PD000001; Euk_pkinase.1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TIR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFD1063 CRC64;
```

Query Match 25.4%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 2,6e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

```
QY 563 PONGSVTCFPEADQVCACHYDPPFCVACPSGVKPLSYMPIWKPFDEGACQCPPI 622
Db 141 PEETATPKTGP--DHCKCAHFLIDGFHCVAACPAGVIGENDTL-VKRYADANAVQQLCHP 197
```

```
QY 623 NCHSCVLDLDDKCPAEQRPASPLTSTIVSAVY-GILLVVLGVVFGILIRROOKIRKRYTM 681
Db 198 NCHSCVLDLDDKCPAEQRPASPLTSTIVSAVY-GILLVVLGVVFGILIRROOKIRKRYTM 253
QY 682 RRLQELTELEPTLPGCAMPNQNOMLLETETLRKTVLGSAGFVGVYGMIPQENIK 741
Db 254 RRLQELTELEPTLPGCAMPNQNOMLLETETLRKTVLGSAGFVGVYGMIPQENIK 313
QY 742 IPVAIKVLRNTSPKANKEILDEAYVAGVSPYRLLGICTSTVQVLTQIMPYGCLL 801
Db 314 IPVAIKVLRNTSPKANKEILDEAYVAGVSPYRLLGICTSTVQVLTQIMPYGCLL 373
QY 802 DHRNENGRISQDLMCMQIKKMSYLEDVVLVARDLAARVVLKSPNHYKITDFGLA 861
Db 374 DHRNENGRISQDLMCMQIKKMSYLEDVVLVARDLAARVVLKSPNHYKITDFGLA 433
QY 862 RLIDIDETEHADGSKVPIKMALESILRRFRTHSDVWSYGVYWEIMTGANPYDGIIP 921
Db 434 RLIDIDETEHADGSKVPIKMALESILRRFRTHSDVWSYGVYWEIMTGANPYDGIIP 493
QY 922 AREIPDLLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFV 981
Db 494 AREIPDLLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFV 553
QY 982 VIO-Q-NEDLGASPLDSTFYRSLLEDMDGDLVDAEEYLVPOQGFPCPDPAFGAGMWHHR 104
Db 554 VIO-Q-NEDLGASPLDSTFYRSLLEDMDGDLVDAEEYLVPOQGFPCPDPAFGAGMWHHR 598
QY 1041 HRSSSTRSGGDLTLGLEPSEEAPSPPL-----APSEAGSDVDFDDLVAGVTGSLQSL 109
Db 599 HRSSSTRSGGDLTLGLEPSEEAPSPPL-----APSEAGSDVDFDDLVAGVTGSLQSL 631
QY 1096 PTHDPSPLQRYSSDPTVPLPSET--DGYVAPLTCSQPPEYVNOPDVRPQPSFREGPLPA 115
Db 632 PTHDPSPLQRYSSDPTVPLPSET--DGYVAPLTCSQPPEYVNOPDVRPQPSFREGPLPA 675
QY 1154 ARPAGATLERAKTLSPGKNGVQDVFAFGAGMWHHR 1191
Db 676 ARPAGATLERAKTLSPGKNGVQDVFAFGAGMWHHR 715
```

```
RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; Pubmed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1.
DR Pfam; PF02813; Retro_M.1.
DR ProDom; PD000001; Euk_pkinase.1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TIR; 1.
```

KW Tyrosine-protein kinase.
 FT NON TER. 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;
 Query Match 25.4%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 2,5e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 572 GPEADQVCAHAKYKDPFCVACRCSGVAPDLSYPMKFPDEACQPCPINCNTSGVDL 631
 1 GP--DHCKKCAFIIGHPCHVACRPAVGENDTL-VMKYDAANAVCQLCHPNCRGKGP 57
 Db
 QY 632 DDXGCPAECRASPSTISAVV-GILVVVLGVVFGILIKRQCKIRKYMRRLDQETEL 690
 58 GLEGCP--NGSKTPSIAGVVGGLCLVVGGLGILYLR--HIVKKTLLRLLDQEREL 113
 Db
 QY 691 VEPLTPSGAMPQAOAMRIKETELKRVVLSGAGFGVYKQIMVPEDEANKIPAIKIVLR 750
 114 VEPLTPSGEAPVQALRIKETELKRVVLSGAGFGVYKQIMVPEDEANKIPAIKIVLR 173
 Db
 QY 751 ENTSPKANEILDEAYVMAVGSVYSRLIGICTSTVQVLTQMPYGLLDHVRNRRGR 810
 174 EATSPKANEILDEAYVMAVGSVYSRLIGICTSTVQVLTQMPYGLLDHVRNRRGR 233
 Db
 QY 811 LGSODLLMKQIAKNGSYLEDVRLVHRDLAARVVLKSPHVKITDPGLARLDIDETE 870
 234 IGSQYLLMKVQIAKNGSYLEDVRLVHRDLAARVVLKSPHVKITDPGLARLDIDETE 293
 Db
 QY 871 YHADGKVPKIMMALESILRRRFTHSDVMSYGVTVLMTFGAKPYDGI PAREIPDLLE 930
 294 YHAEKGKVPKIMMALESILRRRFTHSDVMSYGVTVLMTFGAKPYDGI PAREIPDLLE 353
 Db
 QY 931 KGERLPQPPICITIDVYMVMKCMIDSECRPRELVSFSRMAADPQRFVTVQ-NEDLG 989
 354 KGERLPQPPICITIDVYMVMKCMIDSECRPRELVSFSRMAADPQRFVTVQ-NEDLG 413
 Db
 QY 990 PASPLDSTFYRSLLEDDMDGLVDAEELVYVQGFPCPDPAAGAGVHRRHSSSTRSG 1049
 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLPHQGF-----NSPST--- 454
 Db
 QY 1050 GGDLLTGLPESEEPKRSPL-----APSEGAGSVFPGDGLGMAKAGLQSLPHDPSPLQ 1104
 455 -----SRTPLSLSSATSSNNSATCID-----RNGQCHPVRSEDSFYQ 491
 Db
 QY 1105 RYSEDPVPLPSET--DSGVVAFLTCSPOPEYVNPDPVROPSPRESPLPAARPAATLE 1162
 492 RYSSDPTGNFLPEESIDDSGL-----PAPEYVNG--LMPKKPS----- 526
 Db
 QY 1163 RAKTLSPKNGVVKDVF-----AFGAVENPEYL 1191
 527 ----TAWVQNOIYNNISLTAISKLPMDRYQNSHSTAVDNPXYL 566

RESULT 12
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Man D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF318349; AAL55856.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 25.1%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 5,5e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 883 MALESILRRRFTHSDVMSYGVTVLMTFGAKPYDGI PAREIPDLLEKGERLPQPPIC 942
 1 MALESILRRRFTHSDVMSYGVTVLMTFGAKPYDGI PAREIPDLLEKGERLPQPPIC 60
 Db
 QY 943 IDVYIMVKCMIDSECRPRELVSFSRMAADPQRFVTVQ-NEDLGASPLDSTFYRSL 100
 61 IDVYIMVKCMIDSECRPRELVSFSRMAADPQRFVTVQ-NEDLGASPLDSTFYRSL 120
 Db
 QY 1003 LEDDDMDGLVDAEELVYVQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 106
 121 LEDDDMDGLVDAEELVYVQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 180
 Db
 QY 1063 EAPSPPLAPSGASDVFDGDLGMAAKGIGLSLPHDPSPLQYSSDPTVPLPSETDGYV 112
 181 EAPSPPLAPSGASDVFDGDLGMAAKGIGLSLPHDPSPLQYSSDPTVPLPSETDGYV 240
 Db
 QY 1123 APLTCSPOPEYVNPDPVROPSPRESPLPAARPAATLERAKTLSPKNGVVKDVFARF 118
 241 APLTCSPOPEYVNPDPVROPSPRESPLPAARPAATLERAKTLSPKNGVVKDVFARF 300
 Db
 QY 1183 GAVENPEYLTPQGAAPQ-----HPPA---FSPAFLNL 121
 301 GAVENPEYLTPQGAAPQ-----HPPA---FSPAFLNL 360
 Db
 QY 1215 YVMD-QDPPER-----GAPSTFGCTTAN 1239
 361 WWTTCQCEPBGQVRRSPDVSSGSRGLTSAGIKRMGEPPTTSKGTCHARN 410

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-V-erb-A-V-erb-B protein.
 GN GAG-V-ERB-A-V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11861;
 RN
 RP SEQUENCE FROM N.A.
 RA MEDLINE=30206603; Pubmed=1969616;
 RA Bruskini A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B-";
 RL Oncogene 5:15-24 (1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -
 DR EMBL; X52211; CA36459.1; JOINED.
 DR HSSP; P10828; NTL.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000536; Hormone_rec_119.

DR InterPro: IPR001723; StcHrmn_receptor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR001628; Znf_Cstereoid.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF00105; ZF-C4; 1.
 DR PRINTS: PR00398; STRDHOMNER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR ProDom: PD000035; Znf_Cstereoid; 1.
 DR SMART: SM00430; HOL1; 1.
 DR SMART: SM00219; TykC; 1.
 DR SMART: SM00399; Znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.4%; Score 1653.5; DB 15; Length 962;

Best Local Similarity 51.7%; Pred. No. 4.2e-116; Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

535 VEECRVLQGLPRE-YVNAH-HCLP-----CHPECO 562
 354 IEKCESYLLAFEHYINRKINIPHEWSKLMKVADLRMIGAVHASFHLMKVECPTELS 413
 563 PONGSVTCFGEADQCAAHYKDPFCVAPCPGSGVPDLSTYPIKWFPEBACQPCPI 622
 414 PQE-----VGP--DHCKMCAHFIDGPHCVACPGVIGENDTL-VWKYADANAVCOLCHP 465
 623 NCHSCVDLDKCGPAEQASPLTISVSAV--GILLVVLGVVFGILLIRKROOKIRKYM 681
 466 NCTRGCKGPGLEGCP---NSKTPSIAAGVAGLLCLVVGGLGIGLYLR--HIVKRTL 521
 682 RLLIQTETLVLP--PPSGAMPNQAQMLIKETELRKVKVLCGAGFVYKGIWIPDENYK 741
 522 RLLQERELVEPLTPSGAPNQAHLRIKTEPFKVKVLPFGAFGVYKGLWIPBEKXT 581
 742 IPVAIKVLRNTSPKANKELIDEAYVAVAGVSPVYSRLIGICTSTVOLTQMLPFGCLL 801
 582 IPVAIKELREKTSKANKELIDEAYVAVASVDNPHVCLLIGICTSTVOLTQMLPFGCLL 641
 802 DHVRENGRLGSGDLLWMCQIAKGSYLEDVRLVREDLAARNVLYKSPNHVKITDFGLA 861
 642 DYIREHKDNIQSQYLWMCVQIAKGNVYLEERHVRHDLAARNVLYKTPQHVKITDFGLA 701
 862 RLLIDETEVHADGGKVPKIMMALESLIRRFTHOSDVMSYGVTVWELMTFGAKPYDGI 921
 702 KOLGADEKEYHABEGKVPKIMMALESLIRIYTHOSDVMSYGVTVWELMTFGSKPYDGI 761
 922 AREIPDLLEKGERLPQPICTIDVYIMVWCMWIDSECRPRELSEFSKMAADPQRFV 981
 762 ASEISSVLEKGERLPQPICTIDVYIMVWCMWISGADSRKPFRELIASFKAARDPRL 821
 982 VIO-NEDLGPASPLDSTFYRSLLEDMDMGLVDAEELVPOGFFCPDPAAGAGVTHHR 1040
 822 VIQDERMHLPSPTDSFYRSLMEEDMEDIVDAEELVPHGFF----- 866
 1041 HRSSSTRSGGDDLTLGLEPESEEAEPRLAPSEAGSDVFDGLGMAKGLQSLPTHP 1100
 867 -NSPST-----SRPPLSLSLATSN-----NSATKICIDRNGH-- 898
 1101 SPLQRYSEDPTVLPSTTDGVVAFLTSSPPRYVNOQVPRPQSPREBPLPAAPAGAT 1160
 899 -----PVREDFGL-----PAPEYVNO--LMPKPKSTAMVQNOIYVYSIT 936
 1161 -LEPAKTLSPKNGVGVKDVAFGAGAVENPEYL 1191
 937 AISKLPWDSRYQN-----SHETAVDNPEYL 961

RESULT 14

ID 085468 PRELIMINARY; PRT; 545 AA.

AC 085468;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian Erythroblastosis virus (T834) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N. A.

RX MEDLINE=88217326; PubMed=2697102;
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL: X06943; CA330024.1; -
 DR HSSP: P11362; TRGX

DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase; 1.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; TykC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCAF08AF4 CRC64;

Query Match 24.3%; Score 1645; DB 15; Length 545;

Best Local Similarity 54.9%; Pred. No. 7.8e-116; Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

572 GPEADQCVACAHYKDPFCVAPCPGSGVPDLSTYPIKWFPEBACQPCPINTHSCVDL 631
 1 GP--DHCKMCAHFIDGPHCVACPGVIGENDTL-VWKYADANAVCOLCHPNTCRGKGP 57
 632 DDKGPAPQASPLTISVSAV--GILLVVLGVVFGILLIRKROOKIRKYMRLLOETEL 690
 58 GLEBCP--NSKTPSIAAGVAGLLCLVVGGLGIGLYLR--HIVKRTLRELLQEREL 113
 691 VEPITPSGAMPNQAQMLIKETELRKVKVLCGAGFVYKGIWIPGENVKIPVAIKVLR 750
 114 VEPITPSGAMPNQAHLRIKTEPFKVKVLPFGAFGVYKGLWIPBEKXTIPVAIKELR 173
 751 ENTSPKANKELIDEAYVAVAGVSPVYSRLIGICTSTVOLTQMLPYGCLLDHVRENGR 810
 174 EATSPKANKELIDEAYVAVASVDNPHVCLLIGICTSTVOLTQMLPYGCLLDVIREHKD 233
 811 IGSQDLLNMVCMQIAKGSYLEDVRLVREDLAARNVLYKSPNHVKITDFGLARLLDTEFE 870
 234 IGSYILLNMVCMQIAKGNVYLEERHVRHDLAARNVLYKTPQDKITDFGLAKGADEKE 293
 871 YHADGKVPKIMMALESLIRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAPEIPDLLE 930
 294 YHABGKVPKIMMALESLIRIYTHOSDVMSYGVTVWELMTFGSKPYDGI PAPEISSVLE 353
 931 KGERLPQPICTIDVYIMVWCMWIDSECRPRELSEFSKMAADPQRFVYIO-NEDLG 989
 354 KGERLPQPICTIDVYIMVWCMWISGADSRKPFRELIASFKAARDPRLVYIOQDERM 413
 990 PASPLDSTFYRSLLEDMDMGLVDAEELVPOGFFCPDPAAGAGVTHHRSSSTRSG 104
 414 LPSPTDSKPYRSLMEEDMEDIVDAEELVPHGFF-----NSPST--- 454
 1050 GGDITLGLSEBEARSPV-----APSEAGSDVFDGLGMAKGLQSLPTHPSPLO 110
 455 -----SRPPLSLSLATSNNSATNCIDRNGG-----H----- 481

QY 1105 RYSEDPVPLPSETDGYVAPLTCSPQPEYVNAQPPVPPSPREGPFAAPAGAT-LEP 1163
DB 482 -----PVRBGLF-----PAPRYVQ--LMPKKEPSTANVQIQINYISLTAISK 523
QY 1164 AKTSPGXGVVQVAFAGGAVENPEYL 1191
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter U.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ AND 129/SVEVTAJ;
RA Reiter U.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
MEDLINE=21085660; PubMed=11217851;
RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batolov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barz H., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -
DR EMBL; AF275366; AAG28047.1; -
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23668.1; -
DR EMBL; AK004883; BAB23641.1; -

DR EMBL; AK004911; BAB23662.1; -
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin_Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SMO0261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;
Query Match 22.2%; Score 1503.5; DB 11; Length 655;
Best Local Similarity 43.9%; Pred. No. 5,2e-105;
Matches 281; Conservative 99; Mismatches 229; Indels 31; Gaps 9
QY 11 LLLALLPFGA--STGYC-----TGYIANKKFGITELLVGGQGVVQGNLELYLPT 61
DU 14 LITLCAAGALEKKKQCGTSNLTGLGFEDHFLSL-GRMYNCEVVLGNLEITYVOR 72
QY 62 NASLFLQDIQEVQGVVLIANNQVQVPLRLRIVRGTOULFEDNVALAVLNDGPLNNTT 121
DB 73 NYDSLFLKTIQEVAGVYLALNTVERIPLENLQIRGNALYENTVALALISN----- 124
QY 122 PVTGASPGRLRELORELTLELKGCVLIQRNPOLCYODITLMDI----FKKNQLALTL 177
DB 125 --YGNRTGRELPMRWLOELILGAVRFSNNPILCNMDITQWRDIVQNVFMSMSMDL-- 180
QY 178 IDTRSPRACHPCSPMCKGRSGRWGSSSEDCSLTRTVCAAGCA-RCKGPLPTDCBQCAA 236
DB 181 --QGHFSSCPKCDSCGNSGCGWGEENCKLKIICQAQCSIRCKRGSRSDCHQCAA 238
QY 237 GCTGPKNSDCLALHPNHSIGICELHCPALVTYNTDPESNPNEGRYTGASCVTAQPN 296
DB 239 GCTGPRESDCLVCCQFODEATCTDCTPLLVPTTYQMVNVEGKYSFGATCVKCKCPEN 298
QY 297 YLSTVDVSGCTLVCPLIHQEYTAEDGTORCEKSPCARVYVGMHELEVRVAVTANIQ 356
DB 299 YVYTDHSGCVRACGPDYREV-EDDGIRKCKCKGCPKRVKCVNGIGIEFKDTISNATNIK 357
QY 357 EFAQCKKIFGSLAFIPSSFDGDPASNTAPLPQBPOLQVFTLEITGYLYISAWPDSLPL 416
DB 358 HFKYCTAISGLHILPVAFKGDSFTTRPPLDPRELEITLKVKETITFLLIQAMPDWTDL 417
QY 417 SVFQNLQVITGRILHNAVSLTIOGLGISLVLRSLRELGSGLALHNHNLCLFVTVTW 476
DB 418 HAFENLEIRKTRQHQFSLAVAGLNTISLGRSLKEISDGVIIISGRNLCYANTTW 477
QY 477 DQLEFRNPHQALLHTANPEDECVGEGIACHQLCARGHCWGPCTQCNGNSQFLRQGEVE 536
DB 478 KKLFGTGNQTKTKMNNBAEKDKAVNVHVCNPLCSSGCGWGPBRDCVSGQNVSRGREVE 537
QY 537 ECRVLQGLPREYVNAARCLFCHEGCPQNGSVTCFGEADQCVAAHYDPPFCVARCS 596
DB 538 KCNLLEBEPREFENSSCICQHECCIPQANNITICTRGDPNCTQCAHYIDGRVCVTCBA 597
QY 597 GVKPDLSTYMPIMKPEDEGACQCPINCTHSCTVDLDDKCG 636
DB 598 GIMGEENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCG 636

Search completed: July 22, 2003, 08:59:55
Job time : 53.1263 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:12:49 ; Search time 10.2304 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MELALCRWGLLLALPPGA.....TRKGTPTAENPEYLGLDPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	6729	98.8	1255	1	ERB2_HUMAN
2	5925	87.0	1257	1	ERB2_RAT
3	5918.5	86.9	1254	1	ERB2_MESAU
4	3115	45.7	1210	1	EGFR_HUMAN
5	3094	45.4	1210	1	EGFR_MOUSE
6	2950.5	43.3	1308	1	ERB4_HUMAN
7	2931	43.0	1308	1	ERB4_RAT
8	2669.5	39.2	1167	1	XMRK_XIPMA
9	2398.5	35.2	1342	1	ERB3_HUMAN
10	2332.5	34.3	1339	1	ERB3_RAT
11	1932	28.4	1426	1	EGFR_DROME
12	1749.5	25.7	634	1	ERBB_ALV
13	1703	25.0	604	1	ERBB_AVIER
14	1630	23.9	540	1	ERBB_AVIER
15	1586	23.3	703	1	EGFR_CHICK
16	1275	18.7	1323	1	LIZ23_CAEBL
17	1142.5	18.8	245	1	ERB2_MOUSE
18	742	10.9	1363	1	ILPR_BRALA
19	715	10.5	1382	1	INSR_HUMAN
20	710	10.4	1383	1	INSR_RAT
21	709.5	10.4	1372	1	INSR_MOUSE
22	703	10.3	1300	1	IRR_MOUSE
23	700	10.3	1607	1	MYR_LYMT
24	694	10.2	1297	1	IRR_HUMAN
25	689.5	10.1	1300	1	IRR_CAVPO
26	672	9.9	1477	1	HTK7_HYDAT
27	653	9.6	1367	1	IGLR_HUMAN
28	643	9.4	1373	1	IGLR_MOUSE
29	639.5	9.4	1370	1	IGLR_RAT
30	635	9.3	1390	1	INSR_AEDAE
31	613	9.0	2146	1	INSR_DROME
32	610	9.0	987	1	EPB4_HUMAN
33	591.5	8.7	984	1	EPB1_CHICK

34	589.5	8.7	977	1	EPB2_MOUSE
35	588	8.6	1114	1	RET_HUMAN
36	566	8.6	987	1	EPB4_MOUSE
37	584.5	8.6	976	1	EPB2_HUMAN
38	583.5	8.6	984	1	EPB1_RAT
39	579	8.5	902	1	EPBB_XENLA
40	577.5	8.5	984	1	EPBA_HUMAN
41	576.5	8.5	985	1	EPBA_XENLA
42	573.5	8.4	1053	1	FAK1_CHICK
43	569	8.4	1068	1	FAK1_XENLA
44	567	8.3	757	1	HTI6_HYDAT
45	563	8.3	1052	1	FAK1_MOUSE

RESULT 1	ALIGNMENTS
ERB2_HUMAN	STANDARD: PRT; 1255 AA.
ID	
AC	P04626;
DT	13-AUG-1987 (Rel. 05, Created)
DT	13-AUG-1987 (Rel. 05, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
GN	ERBB2 OR HER2 OR NGL OR NEU.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID:9606;
RN	(1)
RP	SEQUENCE FROM N.A.
RX	MEDLINE=86118663; PubMed=3003577;
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA	Saito T., Toyoshima K.,
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to
RT	epidermal growth factor receptor.";
RL	Nature 319:230-234(1986).
RN	(2)
RP	SEQUENCE FROM N.A.
RX	MEDLINE=86016729; PubMed=2999974;
RA	Consensus L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA	McGrath J., Seeburg P.H., Ullrich A., Schlessinger J.,
RA	Franchete U., Levinson A., Ullrich A.,
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor
RT	shares chromosomal location with neu oncogene.";
RL	Science 230:1132-1139(1985).
RN	(3)
RP	SEQUENCE OF 737-1031 FROM N.A.
RX	MEDLINE=86016729; PubMed=2999974;
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT	human salivary gland adenocarcinoma.";
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN	(4)
RP	VARIANTS VAL-654 AND VAL-655.
RX	MEDLINE=93194196; PubMed=8095488;
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT	"Characterization of a new allele of the human ERBB2 gene by allele-
RT	specific competition hybridization.";
RL	Genomics 15:426-429(1993).
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC	AUTHORITARY NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC	ALPHA AND AMPHIREGULIN.
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC	tyrosine phosphate.
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC	(POTENTIAL).
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11767; AAA35808.1; JOINED.
 DR EMBL; M11761; AAA35808.1; JOINED.
 DR EMBL; M11762; AAA35808.1; JOINED.
 DR EMBL; M11763; AAA35808.1; JOINED.
 DR EMBL; M11764; AAA35808.1; JOINED.
 DR EMBL; M11765; AAA35808.1; JOINED.
 DR EMBL; M11766; AAA35808.1; JOINED.
 DR EMBL; M11730; AAA75493.1; -
 DR EMBL; M12036; AAA35978.1; -
 DR EMBL; X03363; CAA27060.1; -
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; P11362; 1FGK.
 DR Genew; HGNC:3430; ERBB2.
 DR MIM; 164870; -
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; Fu_3.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Polymorphism.
 FT SIGNAL 1 21
 FT CHAIN 22 1255
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1255
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT ACT_SITE 195 204
 FT DISULFID 199 212
 FT DISULFID 220 227
 FT DISULFID 224 235
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 560 560

FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1248 1248
 FT CARBOHYD 68 68
 FT CARBOHYD 124 124
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 654 654
 FT VARIANT 654 654
 FT VARIANT 655 655
 FT CONFLICT 1170 1170
 FT SEQUENCE 1255 AA; 137909 MW; 39BDPDA04DC9362 CRC64;
 SQ
 Query Match 98.8%; Score 6729; DB 1; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MELAIACRMGLLALLPQGAASVTCCTGDMKRLPASPTHLMRLHLYOGCCVQGNL 60
 DB 1 MELAIACRMGLLALLPQGAASVTCCTGDMKRLPASPTHLMRLHLYOGCCVQGNL 60
 QY 61 ELTYLPTNASTLFLDDIOEVQCYIKANSKFIGITELQRLIRVGTOLFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASTLFLDDIOEVQCYVLIANQVRQVPLQRLIRVGTOLFEDNYALAVLDNG 120
 QY 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 DB 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 QY 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 DB 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 QY 181 LTLIDNRSRACHPSPCKSGRSGESSEDCQSLTRFVAGAGCARCGPLPTCCHEQC 240
 DB 181 LTLIDNRSRACHPSPCKSGRSGESSEDCQSLTRFVAGAGCARCGPLPTCCHEQC 240
 QY 181 LTLIDNRSRACHPSPCKSGRSGESSEDCQSLTRFVAGAGCARCGPLPTCCHEQC 240
 DB 181 LTLIDNRSRACHPSPCKSGRSGESSEDCQSLTRFVAGAGCARCGPLPTCCHEQC 240
 QY 241 AAGCTGPKASDCLALFHNSGICELCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKASDCLALFHNSGICELCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
 QY 241 AAGCTGPKASDCLALFHNSGICELCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKASDCLALFHNSGICELCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
 QY 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLREAVAVTSAN 360
 DB 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLREAVAVTSAN 360
 QY 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLREAVAVTSAN 360
 DB 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLREAVAVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPELOVFTLEITGYLYISAMPDLP 420
 DB 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPELOVFTLEITGYLYISAMPDLP 420
 QY 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPELOVFTLEITGYLYISAMPDLP 420
 DB 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPELOVFTLEITGYLYISAMPDLP 420
 QY 421 DISVPCNIOVITGRILHNGAVSLTQIGISMLGRISREAGSGLALHNHTHLCFPHTV 480
 DB 421 DISVPCNIOVITGRILHNGAVSLTQIGISMLGRISREAGSGLALHNHTHLCFPHTV 480
 QY 421 DISVPCNIOVITGRILHNGAVSLTQIGISMLGRISREAGSGLALHNHTHLCFPHTV 480
 DB 421 DISVPCNIOVITGRILHNGAVSLTQIGISMLGRISREAGSGLALHNHTHLCFPHTV 480
 QY 481 PNDQLFRRNHQALLHTNARPEDECEVGEGLAQHQLCARGHCGPGTQCVNSQFLRQEC 540
 DB 481 PNDQLFRRNHQALLHTNARPEDECEVGEGLAQHQLCARGHCGPGTQCVNSQFLRQEC 540
 QY 481 PNDQLFRRNHQALLHTNARPEDECEVGEGLAQHQLCARGHCGPGTQCVNSQFLRQEC 540
 DB 481 PNDQLFRRNHQALLHTNARPEDECEVGEGLAQHQLCARGHCGPGTQCVNSQFLRQEC 540
 QY 541 VECRCYLOGLPREYNNARCLPCHECCOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 DB 541 VECRCYLOGLPREYNNARCLPCHECCOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 QY 541 VECRCYLOGLPREYNNARCLPCHECCOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 DB 541 VECRCYLOGLPREYNNARCLPCHECCOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 QY 601 PGVCKDLSYPMIMKFPBEGACQCPINCCHSCYVDLDDKCPAEPASPLTSVSAVVG 660
 DB 601 PGVCKDLSYPMIMKFPBEGACQCPINCCHSCYVDLDDKCPAEPASPLTSVSAVVG 660
 QY 601 PGVCKDLSYPMIMKFPBEGACQCPINCCHSCYVDLDDKCPAEPASPLTSVSAVVG 660
 DB 601 PGVCKDLSYPMIMKFPBEGACQCPINCCHSCYVDLDDKCPAEPASPLTSVSAVVG 660
 QY 661 ILLVVLGVVFSILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRILKETEL 720
 DB 661 ILLVVLGVVFSILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRILKETEL 720


```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

```

```

Query Match 87.0%; Score 5925; DB 1; Length 1257;
Best Local Similarity 87.0%; Pred. No. 9,7e-309;
Matches 1094; Conservative 49; Mismatches 112; Indels 2; Gaps 2;

```

```

QY 1 MELALGRWGLLALPPGAASITQVCTGIDMKLRLPASPEHLDMLRHYOGGQVVGNTL 60
DB 1 MELAMCRWGLLALPPGIAAGTQVCTGIDMKLRLPASPEHLDMLRHYOGGQVVGNTL 60
QY 61 ELTYLPTRASLSTFODTQEVQOYIKANSKIGITTELORLATYVGTOLFEDNIALAVLDNG 120
DB 61 ELTYVPANASLSFQDQEVQGYVLIAHNVKRPVQRLATYVGTOLFEDKVALAVLDNR 120
QY 121 DPLNNTPTVT-GASPGGLRELOLRSLTEILKGGVLIQBNPOLCYODTILMKDIFHKNNOL 179
DB 121 DPONVAVASTPTGRPEBELREQLRSLTEILKGGVLIQBNPOLCYODMVLKQVFRKNQOL 180
QY 180 ALTLIDTNRSHACHPCSPWCKGSRGWGESSEDCQSLTRIVACAGCARCKPPLPTDCHEQ 239
DB 181 APVIDITNRSHACPPCAPACKDNHCWGESPEDCQILITGICSGCARCKGRPLPTDCHEQ 240
QY 240 CAAGCTGPKHSDDCLACHFNHSGICEHCALTYNTDFEESHPNREGYTGAACVTAC 299
DB 241 CAAGCTGPKHSDDCLACHFNHSGICEHCALTYNTDFEESHPNREGYTGAACVTAC 300
QY 300 PNYLSTDVGSCTLVCPAHNOEVTAEEDGTORCEKSKPCARVCYGLGMEHLREVAVTSA 359
DB 301 PNYLSTEVGSCTLVCPAHNOEVTAEEDGTORCEKSKPCARVCYGLGMEHLREVAVTSD 360
QY 360 NIOEFAGCKIFGSLAPESPFGDPSANAPLOPOLQVFEHLEITGYLITISAMPDL 419
DB 361 NVQEFDCCKIFGSLAPESPFGDPSGIAPLRPOLQVFEHLEITGYLITISAMPDL 420
QY 420 PDLSPQNLQVIRGRILHNGAYSLTQGLISWLGRLRELGSGALIHNTHTLCFVHT 479
DB 421 RDLSPQNLQVIRGRILHNGAYSLTQGLISWLGRLRELGSGALIHNTHTLCFVHT 480
QY 480 VPMQDLFRNHQALHIANRPEDE-CYGEGLACHQLCARGHCKGPPPTCCVNCOSQFLRQ 538
DB 481 VPMQDLFRNHQALHIANRPEDE-CYGEGLACHQLCARGHCKGPPPTCCVNCOSQFLRQ 540
QY 539 ECVEECRVLOGLPREYVNAHCLPCHPECCOPONGSYTCFGEPAADQVCAHAKDPFCVA 598
DB 541 ECVEECRVLOGLPREYVNAHCLPCHPECCOPONGSYTCFGEPAADQVCAHAKDPFCVA 600
QY 599 RCPSGVKRPDLSTYPIKWFPEDEGACQPCPINCTHSCVDLDDKCPAEPASPLTISAV 658
DB 601 RCPSGVKRPDLSTYPIKWFPEDEGACQPCPINCTHSCVDLDDKCPAEPASPLTISAV 660
QY 659 VGLLVVAVGVFGLILKRRQOKIRKTYMRILQETELVEPLTPSGAMNQOMILKET 718
DB 661 VGLLVVAVGVFGLILKRRQOKIRKTYMRILQETELVEPLTPSGAMNQOMILKET 720
QY 719 ELRKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAAYMAGVG 778
DB 721 ELRKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAAYMAGVG 780
QY 779 -SPYVSRLLGICLSTVGLVTLQMLPQYGLDHNVENGRGLSGQDLNWCQIAKMSYLED 838

```

```

DB 781 SPYVSRLLGICLSTVGLVTLQMLPQYGLDHNVENGRGLSGQDLNWCQIAKMSYLED 840
QY 839 VRLVHRDLAARNVLVSPNHVKTDFGLARLLDIDETEVHADGKVPYKMALESILRRR 898
DB 841 VRLVHRDLAARNVLVSPNHVKTDFGLARLLDIDETEVHADGKVPYKMALESILRRR 900
QY 899 FTTHSDVWSGVVWELMTFGAKPYGIPARELPDLLEGERLPQPCITDVMYIMVNC 958
DB 901 FTTHSDVWSGVVWELMTFGAKPYGIPARELPDLLEGERLPQPCITDVMYIMVNC 960
QY 959 WMTIDSECRPFRELVSFMSMARDPFVVIQNEIDGAPASPLDSTFYRSLLEDDMGDLV 101
DB 961 WMTIDSECRPFRELVSFMSMARDPFVVIQNEIDGAPASPLDSTFYRSLLEDDMGDLV 102
QY 1019 DAERYLVPOGCFPCPDPAAGAMVHRRSSSTRGCGGDLTIGLEPSEEAERSLAAS 107
DB 1021 DAERYLVPOGCFPCPDPAAGAMVHRRSSSTRGCGGDLTIGLEPSEEAERSLAAS 108
QY 1079 EGAGSDVFDGDLGGAAGLQSLPTHDPSPLQRYSEDPTVPLPSESTDGYVAPLTCSPQE 113
DB 1081 EGAGSDVFDGDLGGAAGLQSLPTHDPSPLQRYSEDPTVPLPSESTDGYVAPLTCSPQE 114
QY 1139 YVNGPDYRPPPSREGPLPAPAPAGATLERAKTSPGXGVYKDVAFAGAVENPEYLT 119
DB 1141 YVNGSEVQPPPLTPREGPLPAPAPAGATLERAKTSPGXGVYKDVAFAGAVENPEYLT 120
QY 1199 POGGAAPQHPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 1255
DB 1201 POGGAAPQHPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 1257

```

```

RESULT 3
ERR2_MESAU STANDARD; PRT; 1254 AA.
ID AC Q6053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor Protein-tyrosine kinase etbb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
OS ERBB2 OR NEU
NC Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve.
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ueshijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -1- FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL, D16295; BAA03801.1; -.
DR HSBP, P11362; IFGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; Kinase.1.
DR Pfam: PF00757; Furin-like.1.
DR Pfam: PF01030; Recep_L_domain.2.
DR Pfam: PF02757; YLP.2.
DR ProDom: PD000001; Euk_kinase.1.
DR SMART: SM00261; Fu.3.
DR SMART: SM00219; Tyrc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TIR.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1254
FT DOMAIN 22 652
FT TRANSSEM 653 675
FT DOMAIN 676 1254
FT DOMAIN 158 368
FT DOMAIN 472 644
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 236 244
FT DISULFID 240 252
FT DISULFID 255 264
FT DISULFID 268 295
FT DISULFID 299 311
FT DISULFID 315 331
FT DISULFID 334 338
FT DISULFID 511 520
FT DISULFID 515 528
FT DISULFID 531 540
FT DISULFID 544 560
FT DISULFID 563 576
FT DISULFID 567 584
FT DISULFID 587 596
FT DISULFID 600 623
FT DISULFID 626 634
FT DISULFID 630 642
FT MOD_RES 1139 1139
FT MOD_RES 1247 1247
FT CARBOHYD 68 68
FT CARBOHYD 125 125
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 658 658
FT VARIANT 659 659
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C212B2B1 CRC64;

Query Match 86.98; Score 5918.5; DB 1; Length 1254;
Best Local Similarity 86.88; Pred. No. 2.2e-108;
Matches 1089; Conservative 58; Mismatches 107; Indels 1; Gaps 1;

Oy 1 MELALCRMGILLALLPPGASTQVCTGTDMLRLPASPEHLDMLRHLYOCGVVQNL 60
Db 1 MELAAKCGKLLALLSPGASGVCTGTDMLRLPASPEHLDMLRHLYOCGVVQNL 60

Oy 61 ELTYLPNTASLSFLQDIOEVOYIKANSKFIGITEQLRIYRGTOLEFDNYALALNDG 120
61 ELTYLPNTASLSFLQDIOEVOYIKANSKFIGITEQLRIYRGTOLEFDNYALALNDG 120
Db 61 ELTYLPNTASLSFLQDIOEVOYIKANSKFIGITEQLRIYRGTOLEFDNYALALNDG 120
Oy 121 DPLNNTPVYTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILMKDIFKKNOLA 180
121 DPLNNTPVYTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILMKDIFKKNOLA 180
Db 121 DPLNNTPVYTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILMKDIFKKNOLA 180
Oy 181 LTLIDNRSRACHPCSPMKSGSKWSESDOSLTRYCAGGACRCKPLPTDCCHEQC 240
181 LTLIDNRSRACHPCSPMKSGSKWSESDOSLTRYCAGGACRCKPLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGSKWSESDOSLTRYCAGGACRCKPLPTDCCHEQC 240
Oy 241 AAGCTGKHSDDCLACHFNHSGICEHPCALTYNTDTESMPNEGRTYFGASCYTACP 300
241 AAGCTGKHSDDCLACHFNHSGICEHPCALTYNTDTESMPNEGRTYFGASCYTACP 300
Db 241 AAGCTGKHSDDCLACHFNHSGICEHPCALTYNTDTESMPNEGRTYFGASCYTACP 300
Oy 301 YNYLSDVGSCTLVCTLANOEVAEAGTORCEKSKPCARVCTYGLAMEHLREYAVTSAN 360
301 YNYLSDVGSCTLVCTLANOEVAEAGTORCEKSKPCARVCTYGLAMEHLREYAVTSAN 360
Db 301 YNYLSDVGSCTLVCTLANOEVAEAGTORCEKSKPCARVCTYGLAMEHLREYAVTSAN 360
Oy 361 IOEPAGCKKIFGSLAFPLPSFDGDPASNTAPLOPELOVFEETLEITGYLYSAMPDSL 420
361 IOEPAGCKKIFGSLAFPLPSFDGDPASNTAPLOPELOVFEETLEITGYLYSAMPDSL 420
Db 361 IOEPAGCKKIFGSLAFPLPSFDGDPASNTAPLOPELOVFEETLEITGYLYSAMPDSL 420
Oy 421 DLSVPONTQYIRGRILHNGAYSLTLQGLISWLGSLRELSGGLIHNHNLGCVHTV 480
421 DLSVPONTQYIRGRILHNGAYSLTLQGLISWLGSLRELSGGLIHNHNLGCVHTV 480
Db 421 DLSVPONTQYIRGRILHNGAYSLTLQGLISWLGSLRELSGGLIHNHNLGCVHTV 480
Oy 481 PMQOLFENPHQALLHNRNPEDECVESGLACHQLCARGHCMPGPGQCNCQFLRGQC 540
481 PMQOLFENPHQALLHNRNPEDECVESGLACHQLCARGHCMPGPGQCNCQFLRGQC 540
Db 481 PMQOLFENPHQALLHNRNPEDECVESGLACHQLCARGHCMPGPGQCNCQFLRGQC 540
Oy 541 VEECRVLOGLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
541 VEECRVLOGLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
Db 541 VEECRVLOGLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
Oy 541 VKGCRVWKGKLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
541 VKGCRVWKGKLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
Db 541 VKGCRVWKGKLPREYVABHCLPHEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600
Oy 601 PSQVPLSLTNPWKFPDEBGACQPCINCTHSQVLDLKKCPAGORASPLTYSVAVG 660
601 PSQVPLSLTNPWKFPDEBGACQPCINCTHSQVLDLKKCPAGORASPLTYSVAVG 660
Db 601 PSQVPLSLTNPWKFPDEBGACQPCINCTHSQVLDLKKCPAGORASPLTYSVAVG 660
Oy 661 ILLVVLGVVFGILIRPOQKIRKYMRLTETLVEPLTSGAMPNOAQRILKETEL 720
661 ILLVVLGVVFGILIRPOQKIRKYMRLTETLVEPLTSGAMPNOAQRILKETEL 720
Db 661 ILLVVLGVVFGILIRPOQKIRKYMRLTETLVEPLTSGAMPNOAQRILKETEL 720
Oy 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780
721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780
Db 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780
Oy 781 YVERLLGICTSTVOQLVTOQLMPGCLLDHVRERGLSGODLNMVQIAKMSYLEVDR 840
781 YVERLLGICTSTVOQLVTOQLMPGCLLDHVRERGLSGODLNMVQIAKMSYLEVDR 840
Db 781 YVERLLGICTSTVOQLVTOQLMPGCLLDHVRERGLSGODLNMVQIAKMSYLEVDR 840
Oy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHYHADGKVPKIMVALSIRRRFT 900
841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHYHADGKVPKIMVALSIRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHYHADGKVPKIMVALSIRRRFT 900
Oy 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRSLPQPICTIDVYIMVKCM 960
901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRSLPQPICTIDVYIMVKCM 960
Db 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRSLPQPICTIDVYIMVKCM 960
Oy 961 IDSCRRPPELVESESRMAKRDQRFVVIQNEDELGPASPLDSTFYRSLIEDDMGDLVA 102
961 IDSCRRPPELVESESRMAKRDQRFVVIQNEDELGPASPLDSTFYRSLIEDDMGDLVA 102
Db 961 IDSCRRPPELVESESRMAKRDQRFVVIQNEDELGPASPLDSTFYRSLIEDDMGDLVA 102
Oy 1021 EBYLVPQGFCEPDPAAGAGMYHHRHRSSTRSGGDTLGLPSEEBEAPRSLAPSEG 108
1021 EBYLVPQGFCEPDPAAGAGMYHHRHRSSTRSGGDTLGLPSEEBEAPRSLAPSEG 108
Db 1021 EBYLVPQGFCEPDPAAGAGMYHHRHRSSTRSGGDTLGLPSEEBEAPRSLAPSEG 108
Oy 1081 AGSDVDFDGLGMAAGLQSLPHHDSPLQKRSSEDTYVLSBETGTYVAPLTCSPQPEV 114
1081 AGSDVDFDGLGMAAGLQSLPHHDSPLQKRSSEDTYVLSBETGTYVAPLTCSPQPEV 114
Db 1081 AGSDVDFDGLGMAAGLQSLPHHDSPLQKRSSEDTYVLSBETGTYVAPLTCSPQPEV 114

QY 1141 NOPDVPRQPPSPREGULPAPARPGATLEPAKTLSPGKNGVWVDVAFGAGVENEPIYLPQ 1200
 DB 1141 NOPEVRPOPLREGEPLPEVPRPAGATLEPKTLSPGKNGVWVDVTFGGAGVENEPIYLPV 1200
 QY 1201 GCAPCPHPPPSPAFNDLYWDDQPPERGAPSPSTFKGTPAENDEYGLDVPV 1255
 DB 1201 GGSASQPH-PPALCPAFNDLYWDDQPPERGAPSPSTFKGTPAENDEYGLDVPV 1254
 RESULT 4
 EGFR HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD8; Q9UMG5; Q92795; Q00722;
 AC Q00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 Maves E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RA "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=9707686; PubMed=8918811;
 RA Reiter J.L., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Scheinclair C., Pearisall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schein C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=63326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarzky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Cope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=8927817; PubMed=2543678;
 RA Margolis B.L., Lax I., Kriz R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.,
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN (17)
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.,
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN (18)
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.,
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN (19)
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.,
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms, 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -
 DR EMBL; U95089; AAB53063.1; -
 DR EMBL; U48722; AAC50802.1; -
 DR EMBL; U48723; AAC50804.1; -
 DR EMBL; U48724; AAC50796.1; -
 DR EMBL; U48725; AAC50797.1; -
 DR EMBL; U48726; AAC50798.1; -

Query Match 45.7% Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.2% Pred. No. 6.9e-159;
 Matches 622; Conservative 177; Mismatches 360; Indels 106; Gaps 21;

QY 11 LLAALPPGAA--STOVCTGTDMKRLPASPETHLDMLRHLYOCQGVVQGNLLTYLPFN 68
 DB 14 LLAALCPASRLAEKKYQGGTSTNLTQLGTGFEDHFLSLQKMFNKNCEVVLGNLEITTVQRN 73

QY 69 ASISPLDIOEVOYQIKANSKFIETELQRLIRVRCQLFEDNVALVLDNGDPLNNTTP 128
 DB 74 YDISFLKTIQEVAGVYLAALNTEVITPLENIQILRGNNYENSTALAVLSYD----- 126
 QY 129 VTGASPGRLREQLRLSTLEILKGVLYLQNPQLCYODTILWKDIFHKNQALATLIDTNR 188
 DB 127 ---ANKTGLKEIPMNLEQLHGAVRFSNNPALCNVESIQWRDIYSDFLSNMSMDPNH 183
 QY 189 SRAHPFCSMKGSGSCWGESSESDQSILRYCAGCA-RCKGRLPTDCCHQCAAGCTGP 247
 DB 184 LGSQKQDPSCPNPGSCWAGBENCCKLTKIIQAQCCSRCKGKSPSDCHQCAAGCTGP 243
 QY 248 KASDCLACLFHNSICEICELCPALVTYNTDFESPNPEGRTFGASCVTCPYNYLSTD 307
 DB 244 RESDCLVCKKFDKATCDTQCPMLLYPTTYQMDVNPBGKYSFGATVKKCPNNYVTD 303
 QY 308 VSSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYIGMEHLREVAVTSANIQEPAGC 367
 DB 304 HGSQVACGADSYEM-EDGYRCKCKCKGCPKRYCNGIGIGEFKDSLINATNIKHFYNC 362
 QY 368 KXIFGSLAFLESPFGDASNTAPLQEPQLQVETLEITGYLYISAMPDLPPLSVFQN 427
 DB 363 TSIQSDHLPLVAFRGDSFTHTPLDPELDILKTVKETSFLTIQAMPENRTDLHAFEN 422
 QY 428 LQVIRGRLHNGAVSLTQGLGISWLGRLSRLSGSLALIHNTLCFVHTVPMDQLFR 487
 DB 422 LEIIRGRTKQGFSLAVSLNTISLGRSLKEISDGVIIISGNKNCVANTIMWKTLFG 482
 QY 488 NHOQLLHTARPEDEYCEGLACHQLCARHCNGPGPTQVNCQSLRGOECVEEVRVL 547
 DB 483 TSGQTKTIIISNGEASCATQVCHALCSPEGCSPRPDVSGRNYSRGRECVDKNLL 542
 QY 548 QGLPREYNARHCLECPHECQPNQSVYCFEPBADQVACAHYDPPFCVARGSGYKPD 607
 DB 543 EEPREPFENSECIOCHECLPQAMNITCTRGDNQICQAHYIDGHCYKTCAGWNGE 602
 QY 608 LSYPIWKFPBEGACQCPINCTHSCVDLDDKCPAEPORASPLTISVAVGV---ILV 664
 DB 603 NNTL-VKRYADAGHCHLCHNCTYCTGCPLEGCPTNGPKIP--SIATGVGALLLV 659
 QY 665 VLVGVVFGILIKRQOKIRKTYMRLLOETLVLPRLPSGAMPQAOIRLIKETELKVK 724
 DB 660 VALGIG---LNNRRHIVKRLRLRLOERLVLPPLPSGAPQALRLIKETEFKKIK 716
 QY 725 VLGSAGFTVYKGIWIDPGENVKIPVAKYLRNTSPKANKEILDEAYVWAGVSPVYSR 784
 DB 717 VLGSAGFTVYKGIWIDPGENVKIPVAKYLRNTSPKANKEILDEAYVWASVDNPHVCR 776
 QY 785 LLGILSTVOLVQMLMPYGLLDHVENRGLSGQDLNWCQIAKMSYLEDVRLVHR 844
 DB 777 LLGILSTVOLVQMLMPYGLLDHVENRGLSGQDLNWCQIAKMSYLEDVRLVHR 836
 QY 845 DLAAKNVYKSPNNVKTIDFGLARLLIDETEVYADGKVPKIMMALLESILRRRFTHQSD 904
 DB 837 DLAAKNVYKTPQVHKITIDFGLARLLIDETEVYADGKVPKIMMALLESILRRRFTHQSD 896
 QY 905 VMSYGVTVWELMTGAKRPYDGIIPAREIPDLLEKGBRLPOPCTIDVYMTVNCMIDSE 964
 DB 897 VMSYGVTVWELMTGAKRPYDGIIPAREIPDLLEKGBRLPOPCTIDVYMTVNCMIDSE 956
 QY 965 CRPRPRELVSEFSMARSDPQRFVYIQ-NEDLGASPLDSIFYSLSLEDDMGVLVAEY 1023
 DB 957 SRPKRELILFESMAADPQRYVLTQDDEKHLPSPTDSNFYALWDEEDMDVVVADEV 1016
 QY 1024 LVPOQGFPCPDPAFAGAGVHHRRSSSTRSGGADTLTGLPESEEBAPRSLAPSGAGS 1083
 DB 1017 LIPQGF-----SSPSTRKPLSLSSAIS 1042
 QY 1084 DVFGDDLGMAKAKLQGLPHTDSPLQRYSEDPVLPSEST--DGVAAPLTCSPOPEYVN 1141
 DB 1043 N--NSTVACIDRNLQSCPTKEDSFQORYSSDPTGALTEDSIDDTPL-----PVPEYIN 1094

```

QY 1142 QPDRPQPSRPGCUAARAPAGATLFRAKTISPGKNVYKDVAFAGAVENPEYL-TPO 1200
DB 1095 Q-SVPRKPGASVGNPVYHNPPLNP-----APSRDPHYQD--PDRHVGNEPELNTVQ 1143
QY 1201 GGAAPQHPPEPAPFADNLVYWDQ-----DP-----PERGAPPSFTKCTPTAE 1244
DB 1144 -----PTCVNSTFDSAPHAHQSGHSQISIDNDYDQQDFPEAKPNIGFKGS-TAE 1193
QY 1245 NPEYL 1249
DB 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678346;
RA Paria B.C., Das S.K., Andrews G.K., Day S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs W.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Barp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eslinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79897; CAA5587.1; -
DR EMBL; U03425; AAA17899.1; -
DR EMBL; X59698; CAA42219.1; -
DR EMBL; L06864; AAA53029.1; -
DR EMBL; Z12608; CAA78249.1; -
DR HSSP; P11362; 1FGK.
DR MCD; MG195294; EGFR.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TYKc; 1.
DR SMART; SM00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS0011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT NP_BIND 714 981
FT BINDING 720 728
FT BINDING 747 747
FT ACT_SITE 839 839
FT DISULFID 190 199
FT DISULFID 194 207
FT DISULFID 215 223
FT DISULFID 219 231
FT DISULFID 232 240
FT DISULFID 236 248
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333
FT DISULFID 506 515
FT DISULFID 510 523
FT DISULFID 526 535
FT DISULFID 539 555
FT DISULFID 558 571
FT DISULFID 582 591
FT DISULFID 595 617
FT DISULFID 620 636
FT MOD_RES 624 636
FT MOD_RES 680 680
FT MOD_RES 1092 1092

```

```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 (BY SIMILARITY)
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL)
FT CONFLICT 19 19 C -> S (IN REF. 2)
FT CONFLICT 539 539 C -> W (IN REF. 5)
FT CONFLICT 991 991 L -> F (IN REF. 4)
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6)
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2F5 CRC64;

```

Query Match Best Local Similarity 45.4%; Score 3094; DB 1; Length 1210; Pred. No. 9,1e-158; Matches 625; Conservative 170; Mismatches 367; Indels 110; Gaps 23;

```

QY 11 LLLALLPPGAA--STQVCTGTDMKRLPASPERTHLMRLHYOGCOVQGNLELYLPTN 68
DB 14 LITLALCAAGALEKXKVCQGTSTNRILQGLTFEDHFLSLQAMYNCEVGLNEITYVQRN 73
QY 69 ASLSFLQDIOEVQOYIKANSKFGITELQRLIRVGTQLEFEDNYALAVLNDGPDPLNTP 128
DB 74 YDLSEFLKTIQEVAGYVILALNTERIPLENLQIIRGNALYENTYALATLSN----- 124
QY 129 VTGASPGGLRELRLSLTEILKGVLIQNRPOCYODTLIMKDI----PHKNOALALTLI 184
DB 125 -YGNRTGLRELPMKNIQELICAVRFNSNPILCNMDTIQMRDVIQVNFMSWMSMDL--- 180
QY 185 DTRSRACHPCSPWCKSGSRWGBSSBDCSLTRTVAGGCA-RCKGPLPTDCCHQCAAG 243
DB 181 -QSHSPSCPKDCSPSCWGCSENGCQKLTIKICAQCSHRGRGRSDCHQCAAG 239
QY 244 CTGPKHSDCLACHFNHSGICELHCPALTYNTDTFESPNNEGRTTGCACVTACPVNY 303
DB 240 CTGPRESDCLVCKCFDEATCKDTCPPMLYNTTYQMDVNEBGKYSFATCVKKCPRNY 299
QY 304 LSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYCGMEHLEVRAYSANTIOE 363
DB 300 VVTDHGSQVRACGPDYEV-EDDGIKCKKCKGPRKVCNGIGIEFKDTLSINATNIK 358
QY 364 FAGCKKIFGSLAFLEPSFGDPASNTAPLQOPLOVFEETLEITGYLYISAMPDLPDS 423
DB 359 FKYCTAISGDLHLPLPAFKGDSFTTRTPPLDPRELEILTKVKEITGFLLIQAMPDWTDLH 418
QY 424 VFQNLQVIRGRILHNQAVSLTQGLISWLGRLREISGGLALHNHNTLCFVHTVWMD 483
DB 419 AFENLEIIIRGTQHOQPSLAVGLNTISLGRSLKEISDGVIIISGRNLCYANTINWK 478
QY 484 QLFRRNPQALLHTANPEDECVGEGGLAQHQLCARHCWGPFGFTQCVNSQFLRQGEVEE 543
DB 479 KLFGTPNCKTKIMNNAEKDCAVNVHVCPLSSSGCWGPBRDVCVSQNVSRGSECEK 538
QY 544 CRVLOGLPREVYNAHRCLEPCEQCPQNSVTCFEPREADQCYACAHYKDPFCVACRPSG 603
DB 539 CNLSEPREFEVNSCICQHPCLPQAMNITCTGRGPNCTCQCHYIDGPHCVKTCBAG 598
QY 604 VKPDLISYMBIKFPEEGACQPCPINCTHSCVDLDDKCPAEQRASPLTSIVSVAVGLL 663
DB 599 IMGENNTL-VMKYADANNVCHLCNANCTYGCAGPQLQGCWVPSPKPSIATIGVGLL 657
QY 664 VVVLGVVFGI-LIKRQCKIRKYMRLLOELVELVEPITSGAMPNQAMEILKETLRK 722
DB 658 FIVV-VALGIGLPMRRHIVKRLIRLLQERELVEPLTPGGEAPNQAHLIRILKETEFK 716
QY 723 VKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRNTSPKANKELDEAYVMAVGSPYV 782

```

```

DC 717 IKVLGSGAFGVYKGIWIPDGENVKIPVAIKELRENTSPKANKELDEAYVMAVGSPYV 776
QY 783 SRLIGICTSTVQVLTQMLPYGCLLDHRENRRLSGQDLLNMQOIAKMSYLEVRLV 842
DB 777 CRLLIGICTSTVQVLTQMLPYGCLLDHRENRRLSGQDLLNMQOIAKMSYLEVRLV 836
QY 843 HRDLAARNVYKSNHAKITDPGLARLDDIETENYADGKVPKMMALSLIRRRFTTHQ 902
DB 837 HRDLAARNVYKTPQHKITDPGLAKLLGAEENYABGKVPKMMALSLIRLYTHQ 896
QY 903 SDVMSYVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPEICTIDVYMIWKCMID 962
DB 897 SDVMSYVTVWELMTFGSKPYDGIAPASDISILEKGERLPQPEICTIDVYMIWKCMID 956
QY 963 SECPRRREIVSEFSRRARDPQRFVYIQ-NEDGPRASPDLSTRYSLDDEDDGDIYVDE 102
DB 957 ADSRPFKRELILLESKYARDPQRYLYIQDERMHLPSPTDSNFRALMDEDMEDVYDAD 101
QY 1022 EYLVPOQGFPCPDPAQAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSPLAPSEGA 108
DB 1017 EYLVPOQGF-----NSPST-----SRTPPLSSLSA 104
QY 1082 GSDVFGDGLCMGAKGLQSLPTDSEPLQRYSEDPVLPSET--DGYVAPLTCSPQPEY 113
DB 1043 TSN---NSTVACINNGSCRYKEDAFLORYSDPDGAVTEDNIDDAFL-----FVPEY 109
QY 1140 VNQPDYAPQPSREBPPLPAARAGATLERATLSPGKGVKVDVAFGCAVENPEYL-T 119
DB 1093 VNQ-SVPRKAGSVQNPVYHNOGLH-----ABGRLLYQN- PHSNAGNPEYLYNT 114
QY 1199 PQGAAPQPPAPPAFPADNLVYMDQ-----DP-----PERGAPSTFKGTPT 124
DB 1142 AQ-----PYCLSSGPNSPALMTQKSHQMSLDNPDYQDFFPKETKPKGIFKQ-PT 119
QY 1243 AENPEYIGLDPV 1254
DB 1192 AENAEYLRVAPR 1203

```

RESULT 6
ERB4 HUMAN
ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
AC Q15103;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERBB4 OR HER4 (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Ploman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
RP SEQUENCE FROM N.A. (ISOFORM JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Eleuterio K., Corfas G., Paul S., Choi C.J., Rio C., Ploman G.D.,
RA Klagsbrun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester";
RL J. Biol. Chem. 272:26761-26768 (1997).

FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPIC	626	648	NPPTSHDCTCYIPWTHGSHLPPDHA -> IGSSIEDCIGLMD (IN ISOFORM JM-B).
SEQ	SEQUENCE	1308 AA;	146807	MM; 5E4AE0985D88761 CRC64;

Query Match	43.3%	Score 2950.5	DB 1	Length 1308
Best Local Similarity	45.0%	Pred. No. 4.5e-150		
Matches	605	Conservative 183	Mismatches 384	Indels 173
QY	9	WGLLLALPPGAA----	STVCTGTDMKRLRPASBETHLDMLRHLGYGCGVYQGNLELTY	64
Db	8	WVWVSLTVAAGTQVPSDSQSCVAGCTENKLSISDLEQYRALRKYENCEVWGNLEITS		67
QY	65	LFPNASTLSDIOIEVQOYIKANSKFIGITELQRLRIYRQQLREEDNYALALVDNGPELN	124	
Db	68	IEHNRDLSTFLRSRETVGTGVALNQFRLPLENRIIRGKLVEDRATALEFLNVRKOG	127	
QY	125	NTTPVTGASPGGLRELOLRSLEILKGVLIQRPOLCYDPTILMKDIFHKNNQLATLLI	184	
Db	128	NF-----GQELGKLNLEILNGVAYVDQNKFLCYATITIMQDI VRRPWPNSLTLV	178	
QY	185	DYTRSRACHPCSEMKCKSGRCSSESSEDQSLTRYACGGC-ARCKGPLPTDCHEQCAAG	243	
Db	179	STGSSGGCCGRCHKSCTG-RCWGPTNHQOQLTRVCAQCDCGRGCVPAVSPCHREBACG	237	
QY	244	CTGPKXSDCIACLHFENSGICEHLCPALVTYNTDFEEMPRDEGRYTFGAGCVACAPYNY	303	
Db	238	CSPKPTDGFACMNPDSACVCTQCPQFFVYNPTTFQLEHNPNAKYITYGACVKKCHNF	297	
QY	304	LSTDVASCTLVOPLNQVEYTAEDGTQCEKSKSPCARCYVLGMEHLREVAVYTSANIQE	363	
Db	298	V-VDSSSCYRACPPSSKMEY-BENGIMKMKPCDICIOPKACDGIIGTGLMSAQTVDSNIDK	355	
QY	364	FACCKXIFGSLAFLESPFPGPASPNTALQEOLOVETLEITGCLYLIAMSPDLPDS	423	
Db	356	FINCTKINNNLFLVTGTHGDPYNAIEAIDDEKLNVEFTVAIEITGFLINOSWPMNMDFG	415	
QY	424	VFQNLQVIRGLIHNGAYSLTQGGIGIMLRLSLRLEIGSLALIHNNTHLCFHYTPMD	483	
Db	416	VFSNLVTIGRVLVYGLSLILIKQGGIISLQFSLKELSAENIYITDNSNLCYHTINMT	475	
QY	484	QLFRNHQALLTANRPDEDCYGEGLACHQUCANQHCKGPEPTQCVNCSQFLRQECVEE	543	
Db	476	TLESTINQIVIRDRKKAENCAEMWCMNHLCSSDGCGPBDQCLSCRFSSGRICIES	535	
QY	544	CRLTQGLPREYVNAHCLPCHDECP-ONGSVTFGEPADQCVACAHYKDPPEFVAVCPG	602	

```

Dh 536 CNIYDGEFFREFENGSCIVCEDPQCEKMEDELITLHGPGPDNCTCKSHFXDGPNCVEKCPD 595
Qy 603 GVKPDLISYMIKMKFPBEBGACQCPCLNCHSCVDLDDKCC-----PAEGRASPL 651
Dh 596 GLOGANSF--IFKAPDRECHPCFPCNCGCGPTSHDCIYPTGSHSTLFQOHAR-TPD 652
Qy 652 TSIVSAVY-GILVVVYLVGVFGLIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 710
Dh 653 --IAAGVIGGLFVLIVGLTFAYVARKSIK-KKRALRFL-ETELVEPLTPSGAPNOA 708
Qy 711 QMRILKTELKRVKVGSGAFGVYKGINIPDGENVKIIVALKVLRNTSPANKELIDE 770
Dh 709 QRLILKTELKRVKVGSGAFGVYKGINIPDGENVKIIVALKVLRNTSPANKELIDE 768
Qy 771 AYYMAGVGSPPYVRLIGICTSTVQVLTQIMPGCLLDHVRNENRGLSGDILLNMCMOJA 830
Dh 769 ALIMASMDHPLVRLILGVCLSPITQVLTQIMPGCLILEVHKKNISGQILLNMCVOJA 828
Qy 831 KMSYLEDEVTLVHRDLAARNVLYKSNHYKITDFGLARLLIDETFYHADGKVPDIKMA 890
Dh 829 KGMWYLEERLTVHRDLAARNVLYKSNHYKITDFGLARLLIDETFYHADGKVPDIKMA 888
Qy 891 LESILRRRFTHSDVMSYGVYVMTLTFGAKPYDGI PAEIPDLLEKGERLPPQPICTID 950
Dh 889 LECIHRKFTHSDVMSYGVYVMTLTFGAKPYDGI PTREIPDLLEKGERLPPQPICTID 948
Qy 951 VYMIWYKCMWIDSECPREELVSEFSRNARDPQRFVYIQNED-LGPASPLDSTFYRSL 1009
Dh 949 VYVWYKCMWIDADSPPKFEELAEFSRNARDPQRFVYIQNED-LGPASPLDSTFYRSL 1008
Qy 1010 EDDMDGLVDAEYLVLPQCGFPQDPAAGGWMHHRHSSSTRGCGDILLGLEPSEE 1069
Dh 1009 DEEDLEDMDAEYLVLP-QAFNIPP-----LYSRAITDNRN-----ELGSHPPAY 1056
Qy 1070 APRS-----PLAF-SEAGSDVDFDGLGAKAKLOS 1100
Dh 1057 TPMGNGQFYVRCGFAEGGVSVYRAPRTSTIPBAVAGAAELFDDSCNGTIRKVA 1116
Qy 1101 LPTHDPSPLQRYSEDTPVLP-----ETNGYVAPLTCSPQRYVQNPQVPPSPR 1153
Dh 1117 PHVQEDSSTQRYSDADTPVAPERSPRGELDEBGYMTPMKDXKQYLPVPE----- 1167
Qy 1154 EGPLPAPAPAGATLEBAKTLSPKNGVNDVAFAGAVENPEYLPQGGAAPOHPPPA- 1212
Dh 1168 ENPFVSR-----KQDDLO-----ALDNPEYHANSNG-----PPAAE 1199
Qy 1213 -----FSPAPDNLYVWDQDPPERGA--PPSTF 1237
Dh 1200 DEYVNEPLYLNTFANTLGRAEYLNKNNILSMPEKAKKAFDNPYMHSLPPRSTLQHPDY 1259
Qy 1238 KGTPT-----AENPEYL 1249
Dh 1260 QEYSTKYFYKQNGRIPVIAENPEYL 1285

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=spinal cord;
RX MEDLINE=57184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frommert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, RETICULIN AND
CC NIAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF041838; AAD08899.1;
CC EMBL: U52531; AAC53051.1;
CC HSSP: P11362; 1FGK.
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC Pfam: PF02757; YLP; 2.
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00261; FU; 4.
CC SWART: SM00219; TYRKC; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP_BIND 718 985 ATP (BY SIMILARITY).
FT BINDING 724 732 ATP (BY SIMILARITY).
FT ACT_SITE 751 751 ATP (BY SIMILARITY).
FT DISULFID 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1168 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CnC64;

```

Query Match 43.0%; Score 2931; DB 1; Length 1308;
 Best Local Similarity 44.8%; Pred. No. 4.9e-149;
 Matches 604; Conservative 190; Mismatches 393; Indels 160; Gaps 28;

```

QY 1 MELA-ALCRMGILT--ALLPFGASTQVCTGTMKRLRPASPEPTHLMRLHYOGGVQV 57
DB 1 MKLATGLMWGSLVLAARTVQPSASQSVCACTENKLSLSLEQYRALRKYEENCEVVM 60
QY 58 GNLELYLPTNASLSFLQDIQEVQYIKANSKFTIGTELORLRIVRGTOLEFEDNVALVL 117
DB 61 GNLEITSIENHRDLSFIRSIREVGYVALNQFRYLPLENLRIIGTKLYEDRYALALF 120
QY 118 DNGDPLNNTTPVTASAGGLRELOLRSLTEILKGVILQGNPOLCYQDPIILMKDIFHKXN 177
DB 121 LNYRKQGNF-----GLOELGELKNLTLEILNGGVYVDQNKFLCYADDTIHMODIVRNM 171
QY 178 QALATLIDTNRSRACHPCSPMKGRSCWGESSEDCOSLTRTVAGAGC-ARCKGFLPTDCC 236
DB 172 PSNMTLVSTIGSSCGRCHKSCTG-RCMGPTENHCQILRTIVCAEQCDGRKCYGPIYSDCC 230
QY 237 HECCAAGCTGPKSHDCLACHFNHSGICELHCPALVTYNTDTRESNPNEGRTYTRASCY 296
DB 231 HRCAGGSGPKDTCACAKNFNDSGACVTQCQPTQVYNNPTTQLEHNHNAKTYGAFV 290
QY 297 TACPYNLTSTVSGCTLVCPAHNQEVTAEDGTORCEKSKPCARVCGYGLMEHLREVAV 356
DB 291 KKCPHNFY-VDSSSCVAPCSKMEV-EENGIMCKRCPTDICKAKDDGIGTGSLSMSAQTV 348

```

```

QY 357 TSANIOEFACKXIFGSLAFLPESFGDPASNTAPQEPQLOVQFETLEETIGLYISAMP 416
DB 349 DSSNIDFINCTINGNLFLVTGIGHDPYNAIDALDPKLNFRVTRITGLNLQTPM 408
QY 417 DSLPDLSPFONLOVINGRIIHNAGVSLTQIGLISWLGHSIRELSGGLALIHNNHLCF 476
DB 409 PNMIDFVFNLTIGRVLVYSLSLILIKQGITSLQOSLKEISAGNIYITDNNLICY 468
QY 477 VHTVPWQLFRNPHQALLHTANRPEDECYGBGLAQHLCARHGCHQCPGPTQCYNCSQFIR 536
DB 469 YHTINMTLTFSYNQRIVRIDNRANCTCAEGVNCNHLCSNDGCPGPDQCLSCRRFR 528
QY 537 GOECVEECRLQGLPREYVVARHCLPCHPECCP-ONGSVTCFPEZDDQCVACAHYDPDF 595
DB 529 GKITIEECNLYDEFEFEENGSI-CVECDQCEKMEGDLTCHGPRGDNCTKSHFIDGN 588
QY 596 CVARCPGVKPDLSYMPWKFPDEBACQPCPINCTHSCVDLDDKQC-----PA 644
DB 589 CVEKCPVOLGANSF--IFKYADQDRCHCHBNCTQCGNGPSSHDCIYVPMTGSHSLPQ 646
QY 645 EQRASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKTMRRLLQETLEVEPLTFS 703
DB 647 HAR-TPL--IAAGVIGGLFTLVMTALTPAVYARKSIK-KKQALRRFL-ETLEVEPLTFS 701
QY 704 GAMPNQAKMILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPXA 763
DB 702 GTANQOQILILKETELRKVKVLGSGAFGVYKGIWIPGETYKIVALKILNETTGPA 761
QY 764 NKEILDAAVMAGVSGSYVSRLLIGLICLTISVQLVLTQMLMYGCLLDHYRNRGRLSGDL 823
DB 762 NVEEMDALMASVDHVLVRLGLVCLSPITQLVLTQMLPHGCLLEVHCHKDNGISQLL 821
QY 824 NMCQIAKMGSYLEDVRLVYRDLAANVLYKSNHNYKIDFGIARLLDIDETEVHADGK 883
DB 822 NMCQIAKMGVYLEDVRLVYRDLAANVLYKSNHNYKIDFGIARLLDIDETEVHADGK 881
QY 884 VPIWMALLESILRRFTQSDVSYGYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 943
DB 882 MPIMKMALECIHRKFTQSDVSYGYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 941
QY 944 PPICTIDVVMIMKCMWIDSECRPRELVSFESRMAPRQPRVVIQNEO-LGPA SPLDS 100
DB 942 PPICTIDVVMIMKCMWIDSECRPRELVSFESRMAPRQPRVVIQNEO-LGPA SPLDS 100
QY 1003 TFYRSLLEDDMDGLVDAEYLVPOQGFPCPD-----APGA 103
DB 1002 KFFQNLDEDELDMDMAEYLVF-QAFNIPPIYTSRTRIDNRSEIGHSPPAYTPMS 106
QY 1040 GGVVHHHRSSSTRSGGDLTLGLESEEARPSPLAPSEGASGVDFDGLMGAAKGLQ 109
DB 1061 GSGFVYDGGFATQCG--NPMPTYTATTSTIPAPVA--QGATAENFDSCCNGLTRKPV 111
QY 1100 SLPTHDSPLORYSEDPVLPFS-----ETDGYAPLTCSPQEPYVQAPVRPQPPSP 115
DB 1116 VPHVQDESSSTRYSADPTVAPAPRNRAELDEGCVMTPMHDKRQGYLNPVE----- 116
QY 1153 REGPLPAPRAGATLEBAKTLPSGKNGVYKDVAFGAGAVENPEYLTPOGGAAPQHPPPA 121
DB 1168 -ENPFVSR-----KNGDLO-----ADNPEYHSASSG-----PPPA 119
QY 1213 -----FSPAFDNLYYMQDPPERGA--PST 123
DB 1199 EDEVNPEPLVNTFTNALGNAEYMKRSLSVPEKAKKALDNDYNNHSLPPRSTLQHPY 125
QY 1237 FKGTPT-----AENPEYL 1249
DB 1259 LOEYSTKYFYKQNGRIRPIVAENPEYL 1285

```

RESULT 8
 XMRK_XIPMA
 ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
 AC F13388;


```

Db 591 PRCHGILGSDTL-IKXVADKNGGQCPCHQNTGCGSGPGISGCGD-IVSHSLANGL 648
Qy 658 VSGILLVVLGVFGILLIKRQOKIRKRYMRLLQETLEVEPLTPSGAMPNQQRILKE 717
Db 649 VSGILITVIALVILLVLRRLRRRIK-RKRITRCLQKEKLEVEPLTPSGQAPNQAFRIKE 707
Qy 718 TELRKAVVLSGSAFGVYKGIWIPDGENYKIPALIVALENSPKRNEKILDEAVYAGV 777
Db 708 TEFRKQVLSGSAFGVYKGIWIPDGENYKIPALIVALENSPKRNEKILDEAVYAGV 767
Qy 778 GSPYVSRLLGICLTSTVQVLTQMPYCGCLLDHYRENGRLSGQDLINMQIATKMSYLE 837
Db 768 DHPHVCRLLGICLTSAVQVLTQMPYCGCLLDYRQHERICGQWILNVCVQIAKAMYLE 827
Qy 838 DVLVHRDLAARVLYKSPHNVKITDGLARLLDIDETRYHADGVKPKMMALESLIR 897
Db 828 ERHLVHRDLAARVLYKSPHNVKITDGLARLLDIDETRYHADGVKPKMMALESLIR 887
Qy 898 RFTHQSDVMSVGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVMIMVK 957
Db 888 TYTHQSDVMSVGTVMELMTFGSKPYDGIIPAKELASVLENGERLPQPICTIEVYMITLK 947
Qy 958 CWMIDSECRPFRELYSESRKARDPQRFVYVIONEDLGPASPLDSTFYRSLIEDDMGDL 1017
Db 948 CWMIDPSSRRPFRELYSESRKARDPQRFVYVIONEDLGPASPLDSTFYRSLIEDDMGDL 1002
Qy 1018 VDAEELYVPOGFPFPPAPAGAGVWHRRSRSSSTRSGGGLTLGLPESEEPAPSPAP 1077
Db 1003 VDAEELYVPOGFPFPPAPAGAGVWHRRSRSSSTRSGGGLTLGLPESEEPAPSPAP 1025
Qy 1078 SEGAGSDVFDGLGMAKAGLQSLPTHDPSPLOVSEDEPTV-PLPSETDGYVAPLTCSPQ 1136
Db 1026 PTGH-----PYRENSITLRNISDPTQNALEKDLGH----- 1056
Qy 1137 PEYVNOQDVAPQP-----PSPRE-----GAP-ARARAGATLEAKLSGKNGVYKD 1183
Db 1057 -EYVNOQGETSSRLSDIYNPNEDLTDGMPVSLSSQAEITNFSPEYLTNNQNSL--- 1112
Qy 1184 VFAFGAVENPEYLTPOGGAAPQPHPPAFSPAFDNLVYWDPPERGAPESTFGCTPTA 1243
Db 1113 PLVSSGSMDDPDY---QAG-----YQAF-----LPQTGALTNGMFLPAA 1150
Qy 1244 ENPEYLG 1250
Db 1151 ENLEYIG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DC 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (BC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Klaus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todor G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Kato M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAIN OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE p85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M43409; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR F1R; A36223; A36223.
DR HSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 20..643
FT TRANSSEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834..834
FT DISULFID 186..194
FT DISULFID 190..202

```



```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDLCMDITDWRDIWRDARIIVKDNDR
SC -> GQPMVPSGLTPOCAPQDWLYLDDPRLTLTASASK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SO SEQUENCE 1342 AA: 148097 MM: 7201E7F66CA374BD CRC64;

```

Query Match 35.2%; Score 2398.5; DB 1; Length 1342;
 Best Local Similarity 40.3%; Pred. No. 1.3e-120;
 Matches 528; Conservative 189; Mismatches 465; Indels 129; Gaps 32;

```

QY 10 GLLLALLPPGAA--STOYCTGTDMKRLPASPTHLDMLHLYOGQVQGNLELTYLPT 67
DB 11 GLFLSLARGSEVNGSAVCPGTNLGSLVTGDAENQYOTLYKLYERCEVVMGNLEIVLGH 70
QY 68 NASLFLQDIOEVOQYIKANSKFIQITELQRLRIVRGTOFENNVALLVNDGDDPLNNT 127
DB 71 NADSLFLQWIREVYGVLVAMNESTPLPLNKLAVNGTQVYDKKALFVM-----LNIYT 125
QY 128 PVTGASPGGLRELOASLTLEILKGVLIQNNPOLCYQDTILMKDIFHKNQALTLIDTN 187
DB 126 ----NSSHARLQRLTLQTLTEILSGVYIEKNDRLCHMDITDWRDIWRD---AEIVMD 178
QY 188 RSRACHPCSPMKCKSRGCGSESSDQSLTIVCAGGC-ARKGPELTDDCHQCAAGCTG 246
DB 179 NGRSCPPCEHYCKG-KCGPGSEDDQTLTKTTCAPQCNHGFSPNNQCHCECGGCGSG 237
QY 247 PKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRITFGASCVTACPNYVLT 306
DB 238 PDDTDCFACHFHNDGACVPCPQPLVYNNKLTQQLPNNHTXQYQGVGVCVASCINFFV-V 296
QY 307 DVGSCITVCPHNLQEVTAEDGTQRCCKSPKCAKVCYVIGLMEHLREVRVANTQEPAG 366
DB 297 DQTSVCVACPPDKNEVD-KNGLMKCEPCCGLCPKCAEGGSG--SRFQTVDSNIDGFVN 353
QY 367 CKKIFGSLAFLPESFDQDPASNTAPLOPEQLQVFETLEIEITGVLYSAPDLSPLSVFQ 426
DB 354 CTKLGNLDFLILGNLDPWHKIRALDPEKLVNFRVREITGLVNLQSPRHHNPFVSF 413
QY 427 NLQVIRGRIILHNAGY-ILTQIGIISLWGLRSIRELSGLALIHNTLHCFVYTVPMQDL 485
DB 414 NLTTIGRSLYNGFSLIMKNLNVTLGFRSLKETISAGRIYISANROLCHYHNSIMTKV 473

```

```

QY 486 FRNPQALLHTA-NRPDECEVGEGLACHOLCARGHCGWPGPTQVCNCSQFLRQCEVEEC 544
DB 474 LRGPTEERLDIKHNRPRDCAVEGKCDPLCSSGGCGPQGLSCIRNYSRGVCVTHC 533
QY 545 RVLGQLPREYVNAHNCIPHEBECOPQNGSVTCTGPEADQCAAHKKDPFCVACRPSGV 604
DB 534 NPLNGEPREPEHAEKSCSHPCECPMGTATCNGSGSDTCAQCAHFRDGHCVSSCPHGV 593
QY 605 KPDSLMPWKFPEDEGACOPCPINCTHSCVDLDDKCPAEGRA-----SPLTSIVAAVG 660
DB 594 LG--AKGIYKTPVQNECRPHCHENTCGSKGELDDCLGTLVLGKTHLTALTYIAG 651
QY 661 ILLVYLVGVFGLIIKRRQOKIR-KYTMRELLQETELVEPRTSGAMPQAOQRILKETE 719
DB 652 --LVVIFPMQLGSLYWRGRRIQKRAMREYLRGESIEPLDPS-EKANVYLARIFEKE 708
QY 720 LRKXVLSGAGFTVYVYGMIPQSENVKIKVALKYLENTPRANKEIIDEAVMAGVS 775
DB 709 LRKXVLSGAGFTVYVYGMIPQSENVKIKVALKYLENTPRANKEIIDEAVMAGVS 766
QY 780 PYVARLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGQDLNNCMQIAKQMSYLEDV 839
DB 769 AHIVRLGLCPGSSGLQVTLQYLPGLSLDHRQHRGALGQULLNMGVOIAKQMYLIEH 828
QY 840 RLVRDILAAKRVLYKSPNHYKITDFGLARLLDDDETETADGSKVPIKMALESILRRRF 899
DB 829 GMVHRNLAARVLLKSPQOVAVDFGADLLPPDDQLLYSEAKTPIKMALESISHFGX 888
QY 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKEGRLEPOPCITIDVYMIWKCX 959
DB 889 THQSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKEGRLEPOPCITIDVYMIWKCX 948
QY 960 MIDSECPRRRELVESEFSRNARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLEDDMGD 101
DB 949 MIDENIPFTEKELANEFTPRAPRDPRLVIVKRES-GPGIAPGEPHGLTNKKULEVELEP 100
QY 1017 LVDAEELVVOQGFQFCDDPARAGAGVHNRHSSSTRSGGDTLTLELP-SESEARSP 107
DB 1008 ELDDLLELLEED-----NLATTLGASALSLVGTLMRPRGQSLL 104
QY 1076 APSEGAGSDVFDGDLGKAAGLQSLPTH-PSPLQRYSEDDPVPPLP-----SETDGVY 112
DB 1049 SPSSGMY-MPMNQGNLGSCEASVSSSECCRPVSLH-----PMRQCLASSESEGVH 110
QY 1129 A-----PLTCSQPF-----YVNPQDVRPOPSPRECP-----L 115
DB 1102 TGSEAELOEKVSMCRSRSRSPRPGDSAYHQRSLTLPVPLSPGLLEEDVAGVYM 116
QY 1158 PAAPFAGATLERAKTLP--GKNQVV-----KQVFAFGAVENPEYLTPOGGAAPQPHP 121
DB 1162 PDLTKCTPSSREGTILSVGLSSVLGTEEBED-----EYETMNRRRHNSP-PHP 121
QY 1211 PAFSPAFNLNYVD-----QDPERCAPSTKGTPTAENPEYL 1249
DB 1213 RPSLELEIGVEYMDVGSLSLSTQSCPLHPVPIMPAGATTDPDEDYEM 1263

```

RESULT 10
 ERB3_RAT
 ID ERB3_RAT STANDARD; PRT; 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase etbB-3 precursor (EC 2.7.1.112)
 DE (C=etbB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=8522190;
 RX Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Kolland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN (2)
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Kolland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuregulin and their putative receptors, Erb2 and
 RT Erb3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997)
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -;
 DR EMBL: U29330; AAC53050.1; -;
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Furin_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1039 AA; 147545 MW; 0AA5F2402BBDFF1E CRC64;
 Query Match 34.3%; Score 2332.5; DB 1; Length 1339;
 Best Local Similarity 40.5%; Pred. No. 4.2e-117;
 Matches 520; Conservative 170; Mismatches 438; Indels 155; Gaps 34
 QY 3 LAALCRGGLLALLPGAA---STQVCTGTDMRLPASPETHLDMRLHYGCGQVYQGN 59
 DB 7 LQVLC---FLSLARGSEMGNGQAVCPGLNLSTGADNGQYLTLYKECEVVMGN 62
 QY 60 LELTYLTNANSLSTFDIQEQVQYITANSKFITGLQLRIYRGTLFEDNYALAVLDN 119
 DB 63 LEIYLTGHNADLSFLQWIRREVTVYVWANNFESVLPNLRVVRGQYVDGKFAIFVM-- 120
 QY 120 GDPPLNNTPTVAGSPGRLRELQRLSTELIKGGVLTQRNPOLCYODTLMKDIPIHNGQL 179
 DB 121 ---LNNVT---NSHALRQKLTQLEITLISGGVYLRKDKLCHMTIMRDLVRR--- 170
 QY 180 ALTIIDNRSRACHPCSPCKSGSRCKGSESDQSILTRIVCAGGC-ARCKGPLPTDCHE 238
 DB 171 GAELIVKNGANGNCPRCHEVCCKG-RCKGPGDDQILTKITICAPQCGRCFGPNQCCHD 229
 QY 239 QCAAGCTGPKHSDCLCLHFNHSGICELHCPALVYNTDTFESMPYRDEGRYFGASCVTA 298
 DB 230 ECAGGCGSPQDIDCFACRFNDSGACVPRCPREPLVYNNKLTFLQLEPFPHTKIOYGVCVAS 289
 QY 299 CPYNYLSTDVSGCTLVCPPLHNOEVTAEQTORCEKSKPCARCVYGLSMENLREYAVTS 358
 DB 290 CPNHFV-VDQTFPCFACRPPDKMEVD-KHGHKMEPCGGLCPKACBEGTSGG-SRYQTVD 345
 QY 359 ANIQEFPGCKKIGSLAFLPESDGPASNTAPLOBOLOVEETLEITGCVLYISAMPSS 418
 DB 346 SNIDGFNCKIKIIGNDPFLTGLVNPWKIPALDEKLNVRVTAETIGYINISWEPH 405
 QY 419 LPDLSPVONLOVTRGRIHLNGAYS-LTLOGLGISWLGSLRELGSGLLIHNNTHLCFV 477
 DB 406 MHNVSVSNNLTITIGRSLVYRGSLIMNGLNLTSLGFRSLKISAGRYVIAANOQLCH 465
 QY 478 HTYPMWDLFPNPHQALLHTA-NRPEDECVGEGLAQQLCARHGCWPGPTQVCNCSQFLR 536
 DB 466 HSLNMTRLLLGPEERLDIKYRFPLECCLAEGVVCPLCSSGCGWPGPGQCLSCRNSR 525
 QY 537 GQECVEECRYLQGLPREYVYARHCLPCHPECOPONGSVTCFPEADQCYACAHYDPPFC 596
 DB 526 EGVCTVHCNLTLOEPREFVHEAQCSCHCECLPMESTISCNSSGSDACRCAHFDGSHC 585
 QY 597 VAPCGVGVKFDLSYPMIKWFPDEEGACQCPINCTHSC--VLDLDDKCGPAGRCASPLTSI 654

```

Db      586 VNSCPHGILG--AKGDIKYKPDANECECHENCTOGCGNPELODCLQAELMSKPHLV
Qy      655 VSAVGLILVWVIGVVFGLIKRPOOKIR-KYTRMLQETELVPLPSGAMPQAMR
Db      644 IANTVG--LAVILMIGSFLVWRGRIONKAMRYLERGESIPLPDS--EKAKVLAR
Qy      714 ILKETELRKVKVIGSAGFTVYKGIWIPDGENVKI PVAIKVLENTSPKANKELIDEAYV
Db      701 IFKETELRKLVIGSAGFTVYKGIWIPDGENVKI PVAIKVLENTSPKANKELIDEAYV
Qy      774 MAGVGSPPVYSLIGITLTSTVQLVQLMVMYGLDHLVRENGRGLSCQLLWMCQIAKGM
Db      761 VGSLDHAAHIVRLGLCPGSSLDQVLYQLPLGLSLDHVQHRETLGPPQLLWVGQIAKGM
Qy      834 SYLEDVRLVHRDLAANVLYKSPNHVKITDFFGLARLIDIDETEVHADGKVPKIMWALS
Db      821 YLLEHSMVHRDLAANVLYKSPNHVKITDFFGLARLIDIDETEVHADGKVPKIMWALS
Qy      894 ILRRRTHSGDVSWSVGVWVWELMTGAKRYDGIIPAREIPDLLEKGERLPPOPICTIDVYM
Db      881 IHFGKTHSGDVSWSVGVWVWELMTGAKRYDGIIPAREIPDLLEKGERLPPOPICTIDVYM
Qy      954 IMVCKMIDSECRPFRELVSEFSRMADPOCFVIVIQNEDLGPASPLDSTFYRLIEDDD
Db      941 VMVKCMIDENIRPTFKELANFTMARDPRYIVIKAS--GPGTP--PAAEPSVLTATKE
Qy      1014 MGDVDAEYLVYPQCGFCPPDPAAGAGVHRRHSSTSGGSDLTGLPSEEE-----
Db      998 L-----QEAELEPEL-----DLIDLEAELEGATS 1023
Qy      1069 -----EAPRPLAPSEG-----AGSDVECDLGMGAAGLOSULTHD 1105
Db      1024 LGSALSTPGTILTRGSGSLSPSSGTPMWNQSLGACLDASVLRGREFSPISLH- 1082
Qy      1106 PSPLORYSEDPVTPVPPSETDGVY-----APL-----TC-----SPOPE-----YVNPQDV 1145
Db      1083 PIPGR-----PASESEGHVTGSEAELOEKVSVCSRSRSPRPROGSAVHSRHS 1135
Qy      1146 RQPPSPREGP-----LPAAKAGATLEAKLSP-GKNGVY-----KDYFAF 1187
Db      1136 LITVTPUSPPELEEDNGVMPPTHURGASSSREGTSLVGLSVLGTDEEDD----- 1191
Qy      1188 GGAVENPEYLTPOGGAPOPHPP 1210
Db      1192 -----EEVEYMRKRGRSP-PRPP 1209

RESULT 11
EGFR_DROME
ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05. Created)
DT 15-DEC-1998 (Rel. 37. Last sequence update)
DT 16-OCT-2001 (Rel. 40. Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DS (Guinea receptor) (torpedo protein) (Drosophila relative of ERBB)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=45124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX MEDLINE=Oregon-R; TISSUE=embryo;
RC STRAIN=87002474; PubMed=3093080;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcript.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amarantides P.G., Scher S.E., Li P.W., Horkins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.C., Chame M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brannan D., Boltsakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu L., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., DeJcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleby J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.G., Wu D., Wiesenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Madeworth S.C., Vincent W.S. III, Bilocheau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

```



```

Db      609 ISNAK- FDNRTKICHPECR-----TCNAGAGADHCQECVHVRDQGVCSBPCPNKYN 660
Qy      602 -----SGVK-----PDLSTYPIW 614
Db      661 DRGVRECHATDCGCTGPKDTIGTACTTCNLAIINNDATVAKGLKDKCPD-GI-FW 717
Qy      615 KF--PDEBGACQ-----CPI-----NCTH-----632
Db      718 EYVHPOEGSLKPLAGRAVCRKCHPLCELTNNYHGEQVCSKTHKRRECECECPADH 777
Qy      633 -----SC-----VLDLXG-----CPAD 646
Db      778 YTDGEQCECFORHECEGCTGPPADDCCKRNKFLPDANETGPPYVSTNCTSKCPLEW 837
Qy      647 R-----ASPLTS-----YSAAVGLLVVVLGVFGILIKRQ 679
Db      838 RHNVYQYTAIGPYCAASPPRSSKITANLDVNMIFITGAVLVPTICLV--TYICRQK 895
Qy      680 QKIRKYT--MRLLQETELVEPLPSGAMNQAMRIKETELRKXVVGSAFGTVYK 737
Db      896 QKAKETVKNMALSCEDESEPLRPSITGNLCKRLIVKDAELRKGVLGMAFGRTVYK 955
Qy      738 IWPDGENVKIPVAIKVLRNTSPKANKELDEAYVAVGSPYVRLGICLTSTVOLV 797
Db      956 VVVEGENVKIPVAIKELKSTGAESESEPLRATYIMASEHNNLLKLAVCMSSQMLI 1015
Qy      798 TQMPYGCILDVHRENGRLGSDLLNMCQIAKMSYLEDVRLVFRDIAANNVLKSPN 857
Db      1016 TQMPYGCILDVYNNNDKIGSKALNWSQIAKMSYEEKLVHARDIAANNVLVQTPS 1075
Qy      858 HVKITDFGLARLDIDETEHADGQKVPFKMALLESIRRFTHSDVMSYVTWELMT 917
Db      1076 LVKITDFGLAKLSSDSNEYKAGKMPKIMLLECIKRRVFTSKADVAFGVTTWELT 1135
Qy      918 FGAKPYGIPAREIPDLLEKGERLPPOPIITDVIMVYKCMWIDSECPREELVSEPS 977
Db      1136 FGQRPHEINPAKIDPLIEVGLKLEQPEICSDIYCTLLSCWMLDAMKPTKQTLTVA 1195
Qy      978 RMARDDPQRFVIONEDLG--PASPLDSTFRLSELD--DMQDLVAEYLVPQGFPC 1032
Db      1196 EFARDPGRYLAIPGDKTRLP-----YTSODEKDLIRKLAFTTDSSEALAKDDYIQ 1248
Qy      1033 PDPAFGAGVYHRRSSSTRSGGDLTLGLEPSEBAP-----RSLPAPSEAGASDV 1086
Db      1249 PKAAPGPS-----HRDCT-----DEMEXLARVCKDPSNKSSSTGDDER 1287
Qy      1087 DG---DLGMAKAGLQSLPHDPSLQRISEDPVTVLSESTGYAPLTCSQPBYVNP 1143
Db      1288 DSSAREVGVGNLR-----IDLVPDEDDYIMP--TCQPGPNNNNM 1325
Qy      1144 DVRRPPSPREPEPLPARAPAGATLERAKTILSPKNGVVCVAFAGAVENPEYL---TP 1199
Db      1326 N-----NPNQNNMAAVGVAAGYM-----DLGVSVSDNPEYLANACTL 1364
Qy      1200 QGGAAPQPH-----PPAPSP-AFDNLVYWD 1224
Db      1365 GVGESPIPTGTTIGIPVWGPGTVEVKVPMGSEPTSSDHEVND 1408

```

```

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
NCBI_Taxid=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA MEDLINE=85228222; PubMed=2988784;
RX Nilsson T.W., Maroney P.A., Goodwin R.G., Rotman F.M.,
RA Crittenden L.B., Raines M.A., King H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INVIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFVLV.
CC HSP; P11362; 1FGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TykKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399
CC NP_BIND 138 146
CC BINDING 165 165
CC ACT_SITE 257 257
CC FT BY SIMILARITY.
CC ACT_SITE 257 257
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 2.2e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 587 CAHYKDPFCVACPSGVKPDLSYPMKFPDEBACQPCPINTCHSCVLDLDDKCPAPQ 646
Db 3 CAHFIDGPHCVACPAVGLGENDTL-VWKYADANAVCQCHNCRGCKGGLBECP--- 58
Qy 647 BASPLTSISAVV-GILVVYGVVFGILIKRQKIKYTKRLLOTELYEPLTPSGA 705
Db 59 NSKSTPSIAGVAVGGILCVAVGLGIGYLRRR-HIVKRTILRLLOREVEPLTPSGE 117
Qy 706 MPNQAMRIKETELRKXVYVLSGAFGVYKGIWIPDENYKIPVAIKVLRNTSPKXN 765
Db 118 APVQHLRLIKTEFEKXVYVLSGAFGVYKGIWIPBEKIKIPVAIKELNKAISPKXN 177
Qy 766 ELIDENYVMAVGSPPVSRLLGICLTSTVOLVLTQMPYGCILDVHRENGRLGSDLLN 825
Db 178 ELIDENYVMAVSDNPHVCRLLGICLTSTVOLVLTQMPYGCILDVIREKNDIGSQYLLNW 237
Qy 826 CMQIAKMSYLEDVRLVFRDIAARVVLKSPHVKITPFGARLDIDETEHADGQKVP 885
Db 238 CQIAKMSYLEDVRLVFRDIAARVVLKTPQHVKITDFGLAKLGADEKGYHAGGKVP 297
Qy 886 IKMALLESILRRRTHSDVMSYVTWELMTFGAKPYGIPAREIPDLLEKGERLP 945

```

```

DB 298 IKMALESLHRIITTHQSDVWSGYATWELMTFGSKRYDGPASELSISVLEKGERLPDP 357
QY 946 ICTIDVYMIWVKCMWIDSECRPRERELVSFSESRMARDPQRFVYIQ-NEDUGPASPDLSTF 1004
DB 358 ICTIDVYMIWVKCMWIDADSRPKRELIAFSSKMARPPRYLVYQGERMHLSPFDSKF 417
QY 1005 YRSLLDDDDMDLDVADEHYLVPOGFCPCPPAPAGAGMWHRRSSSTRGGGDLTLGLE 1064
DB 418 YRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEPAPSP-----APSEGAGSDVDGDLGMAKGLSLPTHPSPFORSEDPYV 1119
DB 450 -----SRTPLLSLSLATSNNSAANCID-----RNGGCHVREDSPFORSSPOTCN 495
QY 1120 LPSET--DGYVAPLTGSPQPEYVNPQVPQPSPREGRLPAARPAATLERAKTISPCK 1177
DB 496 FLEBSIDDGFL-----PAPRYVNO--LMPKKPS-----TAMQ 526
QY 1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQPPAPSPAFDNLVY 1222
DB 527 NQYNNISLTAISKLPMSRYONSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
QY 1223 WDQ-----DPP-----KQAPSTFKGPTAENPEYLGIDVP 1254
DB 579 WIQSGNQIINLDNPDYQDFLPNETKPKVGLKVAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
CN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN 11
RC STRAIN=H;
RP SEQUENCE FROM N.A.
RX MEDLINE=84026539; Pubmed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN 12
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; Pubmed=6328659;
RA Deblut B., Henry C., Benalissa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: K02006; AAA42394.1; ALT_INIT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TYOH.
DR HSPD: P11362; IFGK.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE: PS00111; PROTEIN KINASE DOW; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146
FT BINDING 165 165
FT ACT_SITE 257 257
FT CONFLICT 29 29
FT CONFLICT 140 140
FT CONFLICT 146 146
SQ SEQUENCE 604 AA; 67633 MM; 76EBDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best local similarity 52.2%; Pred. No. 6,4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 587 CAHYKDPFCYARCGSKGKBDLSMPYMKRPFDEGACQPFINCTHSCVLDKGGPAG 646
DB 3 CAHFTDGHCVKACPAVGLGNDTL-VKRYADANAVQCLHPCTGCKGPGLEGCP--- 58
QY 647 RASPLTSIVSAV-GILLVVLGVVGFILIKRQOKIRKTYMRLLQETELVEPLTPSGA 705
DB 59 NGSKTPIIAACVVGGLLVVVGIGLILRR-HYRKXLTLLRLQERLVEPLTPSGE 117
QY 706 MPNQAKILKETELRKVKVLGSAFGTVKGIWIPDGENVKIPLVAILURENTSPYAK 765
DB 118 APQAHARILKETEFKVKVGLGSAFGTVKGIWIPDGENVKIPLVAILURENTSPYAK 177
QY 766 EILDEAYMAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLHVHVENRGLSGODLLM 825
DB 178 EILDEAYMASVDNPHVCRLIGLICLTSTVQVLTQMLPYGCLLHVHVENRGLSGODLLM 237
QY 826 CMQIAKGSYLEDEVLRHDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADGKVP 865
DB 238 CVOIAKGMNYLEERLVRHDLAARNVLKTPQVKTDFGLARLLDIDETEVHADGKVP 297
QY 866 IKMALESLHRRTHQSDVWSGYATWELMTFGSKRYDGPASELSISVLEKGERLPDP 945
DB 298 IKMALESLHRIITTHQSDVWSGYATWELMTFGSKRYDGPASELSISVLEKGERLPDP 357
QY 946 ICTIDVYMIWVKCMWIDSECRPRERELVSFSESRMARDPQRFVYIQ-NEDUGPASPDLSTF 1004
DB 358 ICTIDVYMIWVKCMWIDADSRPKRELIAFSSKMARPPRYLVYQGERMHLSPFDSKF 417
QY 1005 YRSLLDDDDMDLDVADEHYLVPOGFCPCPPAPAGAGMWHRRSSSTRGGGDLTLGLE 1064
DB 418 YRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEPAPSP-----APSEGAGSDVDGDLGMAKGLSLPTHPSPFORSEDPYV 1119
DB 450 -----SRTPLLSLSLATSNNSAANCID-----RNGGCHVREDSPFORSSPOTCN 495
QY 1120 LPSET--DGYVAPLTGSPQPEYVNPQVPQPSPREGRLPAARPAATLERAKTISPCK 1177
DB 496 FLEBSIDDGFL-----PAPRYVNO--LMPKKPS-----TAMQ 524
QY 1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQPPAPSPAFDNLVY 1222
DB 527 NQYNNISLTAISKLPMSRYONSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 574
QY 1219 NLVYWDQDPPERGAPSTFKGPTAENPEY 1248

```

Db 575 S8FYWIGSGNHQ-----INLNDPBY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.

AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.";
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
CC EMBL: M13179; AAA2401.1; .
CC DR PIR: A25231; TVFVEB.
CC DR HSSP: P11362; 1FGK.
CC DR InterPro: IPR000719; Euk_pkinase.
CC DR InterPro: IPR001245; Ty_pkinase.
CC DR Pfam: PF00069; pkinase; 1.
CC DR ProDom: PD000001; Euk_pkinase; 1.
CC DR SMART: SM00219; TyKc; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399 PROTEIN KINASE.
CC FT NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CnC64;

Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 4.4e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 567 CAHYKDPFCVACRCSSGKVPDLSTVPIKPFDEGACQPCPCINCHGCVDDIDKGCFAEQ 646
Db 3 CAHFDGPHCVACAGVAGLGVNDL-VKRYADANAVQCLCPNCTRCKPGLEGCP--- 58
QY 647 RASPLTSIVSAVV-GILVVVVGAVFGILIRROOKTRKYMRRLQETELVEPLTSGA 705
Db 59 NGSKTPTSIAGVVGGLLGLVVGVLGIGLYLR-R-HYVRKRLRLDLERLVEPLTSGE 117
QY 766 MPMNQMRILKETELRKRYV-GSGAFGVYGVIVPDGKNVKAIVAKIVRENTSPANK 765
Db 118 AFNQHLRLILKETELRKRYV-GFGAFGVYGVIVPDGKNVKAIVAKIVRENTSPANK 177
QY 766 EILDEAVYMAVGVSFVYRLLGLCTSTVQVLVLTQMPGCLLDVVRNRRLQSODLNN 825

Db 178 EILDEAVYMAVSDNPHVCRLLGLCTSTVQVLVLTQMPYGCLLDYLREHKDNGISQVLLNW 237
QY 826 CMQIAKMSVYEDVRLVRLAARVVLVKSNNHYKTTDFGLARLLDIDETRYHADGKVP 885
Db 238 CVQIAKGMNVYEEERKHKVRDLAARVVLVKTQHYKTTDFGLAKQGADEKEYHABGGKVP 297
QY 886 IKMMALESILRRRFTHSDVMSYGVTVMLTFGAKPYDGI-PAREIPDLLEKGRLLPQP 945
Db 298 IKMMALESILRRITVHSDVMSYGVTVMLTFGSKPYDGI-PASEISVLEKGRLLPQP 357
QY 946 ICTIDVYIMVAKMIDSECPREFVELSESRKARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVAKMIDSECPREFVELSESRKARDPQRFVVIQ-NEDLGPASPLDSTF 417
QY 1005 YSLDEDDMDGLVDAEYVLPQCGFCPPDAPAGVGHRRHSSSTRSGGDTLGL 1064
Db 418 YRTIMEEEMEDIVADYVLPQCGF-----NSPST----- 449
QY 1065 PSEEBAPNSPL-----APSEGASDVFDGLGMAAKGLSLPHDPSPLGRYSEDPTVP 1115
Db 450 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 476
QY 1120 LPSRDGVVAPITCSPOEYVNOQPVVRQPSRPGPLAARPAAGT-LEKAKTSPCKN 1178
Db 477 -PVREDGFL-----PAEYVNO-LMKKPSSTAMVQNYVISTALISKLPIDSRQN 527
QY 1179 GVYKDVFAFGAVENPEYL 1197
Db 528 -----SHSTAVDPEYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.

AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellor F., Winkler M.,
RA Ulrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.";
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MEALALCRWGLLALLPFGA.....TFKGFPAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671560 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_ricent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6223	91.4	1259	6 O18735	O18735 canis fam1
2	3120	45.8	1209	11 O90X70	O90X70 ratu
3	3091	45.4	1210	11 O9EP98	O9EP98 muscu
4	2729.5	40.1	1137	13 O9W6F6	O9W6F6 gallu
5	2703	39.7	1165	13 O9YH40	O9YH40 xiphop
6	2275	33.4	1328	13 P79754	P79754 fugu
7	2011.5	29.5	1433	5 O9B1H9	O9B1H9 anophe
8	1797.5	26.4	419	4 O9UK79	O9UK79 homo
9	1739	25.5	367	11 O8R2X1	O8R2X1 muscu
10	1720	25.3	729	15 O8E712	O8E712 avian
11	1718	25.2	567	15 O8E714	O8E714 avian
12	1697.5	24.9	412	4 O8WYV0	O8WYV0 homo
13	1653.5	24.3	962	15 O64895	O64895 avian
14	1645	24.2	545	15 O85468	O85468 avian
15	1482.5	21.8	655	11 O9WVF5	O9WVF5 muscu
16	1466.5	21.5	643	11 O9ERV6	O9ERV6 muscu

17	1274	18.7	1193	5 O9Y1X8	O9Y1X8 ephydactia f
18	1173.5	17.2	1368	5 O23821	O23821 caenothabdi
19	1146	16.8	1217	5 O26566	O26566 schistosoma
20	1117	16.4	527	13 O90836	O90836 gallu
21	982.5	14.4	478	11 O9E8E0	O9E8E0 ratu
22	933.5	13.7	599	13 O9PSH2	O9PSH2 gallu
23	906	13.3	165	4 O14256	O14256 homo
24	887	13.0	176	11 O923V5	O923V5 ratu
25	806.5	11.8	345	13 P11776	P11776 xiphop
26	778	11.4	435	5 O8E2H1	O8E2H1 xiphop
27	754.5	11.1	311	13 O9P162	O9P162 xiphop
28	753.5	11.1	1362	13 O9PV24	O9PV24 xenopu
29	732	10.8	1671	5 O9NJV5	O9NJV5 biemphalari
30	723	10.6	149	6 O9B6G6	O9B6G6 oxyctolagus
31	702	10.3	1418	13 O93457	O93457 scopthalmu
32	701.5	10.3	1368	13 O8WMS5	O8WMS5 paralicthy
33	692	10.2	331	4 O9EUD7	O9EUD7 homo
34	678.5	10.0	1369	13 O8WMS6	O8WMS6 paralicthy
35	678	10.0	1358	13 O73798	O73798 xenopu
36	658	9.7	1412	13 O8WMS4	O8WMS4 paralicthy
37	655.5	9.6	1472	5 O9U5A8	O9U5A8 bombyx mori
38	646.5	9.5	1418	13 O8WMS3	O8WMS3 paralicthy
39	645	9.5	1245	13 O9YGH8	O9YGH8 scopthalmu
40	642	9.4	1571	11 O9QVW4	O9QVW4 ratu
41	617.5	9.1	2144	5 O9VD94	O9VD94 drosophila
42	600	8.8	935	4 O9ELJ5	O9ELJ5 homo
43	599	8.8	987	11 O91YMO	O91YMO muscu
44	596	8.8	987	11 O9NMR2	O9NMR2 muscu
45	587.5	8.6	1036	4 O07912	O07912 homo

ALIGNMENTS

RESULT 1

O18735

ID O18735

AC O18735

DT 01-JAN-1998 (Tremblrel. 05, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Erib-2.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "CDNA cloning of erib-2 from canine mammary gland.";

RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB008451; BAA2127.1; -

DR HSSP; P11362; IFGR.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_Pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_Pkinase.

DR InterPro; IPR004019; YLP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Pkinase; 1.

DR Pfam; PF01030; Recep_L_domain; 2.

DR Pfam; PF02575; YLP_2.

DR ProDom; PD000001; Euk_Pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; Tyrc; 1.

DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 91.4%; Score 6223; DB 6; Length 1259;
 Best Local Similarity 91.2%; Pred. No. 0;
 Matches 1149; Conservative 39; Mismatches 66; Indels 6; Gaps 2;

QY 1 MELAALCRWGLLALLPGAASTQVCTGDMKLRJSPASPTHLDMLRHL YGCGQVVGNTL 60
 DB 1 MELAAMCRWGLLALLPGSAGTQVCTGDMKLRJSPASPTHLDMLRHL YGCGQVVGNTL 60

QY 61 ELTYLPTNASLSLQIQEVQOYIKANSKFIGITELQRLRIYRGQLEEDNYALAVDNG 120
 DB 61 ELTYLPTNASLSLQIQEVQOYIKANSKFIGITELQRLRIYRGQLEEDNYALAVDNG 120

QY 121 DPLNNTPTVTSAPGSLREIQLSLTEILKGVLIQRPOLCYQDTIIMKDIFHNQOLA 180
 DB 121 DPLNNTPTVTSAPGSLREIQLSLTEILKGVLIQRPOLCYQDTIIMKDIFHNQOLA 180

QY 121 DPLNNTPTVTSAPGSLREIQLSLTEILKGVLIQRPOLCYQDTIIMKDIFHNQOLA 180
 DB 121 DPLNNTPTVTSAPGSLREIQLSLTEILKGVLIQRPOLCYQDTIIMKDIFHNQOLA 180

QY 181 LTLIDTNRSPACHPGSPMCKSGSKGSESEDCQSLTRYACAGGACARCGPLPTDCCHQC 240
 DB 181 LTLIDTNRSPACHPGSPMCKSGSKGSESEDCQSLTRYACAGGACARCGPLPTDCCHQC 240

QY 241 AAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGCTLVCPFLHNOEVTAEQTCRCKSCSPCARVCGYGMHLEVRATISAN 360
 DB 301 YNYLSTDVSGCTLVCPFLHNOEVTAEQTCRCKSCSPCARVCGYGMHLEVRATISAN 360

QY 361 IOEFAGCKKIFGSLATLPESFDODPASNTAPLOPELOVFEETLEITGLYLSAMPDLP 420
 DB 361 IOEFAGCKKIFGSLATLPESFDODPASNTAPLOPELOVFEETLEITGLYLSAMPDLP 420

QY 421 DLSVFQNLQYIRGRILHNGAYSLTLOGIGISWIGLRSLRELGSGALIHNTLFCVATV 480
 DB 421 DLSVFQNLQYIRGRILHNGAYSLTLOGIGISWIGLRSLRELGSGALIHNTLFCVATV 480

QY 481 PMDQLEFRNPQALHTANRPEDECVGEGLAGHQLCARGCHWCPGPTQCVNCSQFLRGEC 540
 DB 481 PMDQLEFRNPQALHTANRPEDECVGEGLAGHQLCARGCHWCPGPTQCVNCSQFLRGEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDPPEVCARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDPPEVCARC 600

QY 601 PSQVKPDLSTMPYWKPPDEBGAQCPPICTHSCVDLDKXGCPAQORASPLTISAVVG 660
 DB 601 PSQVKPDLSTMPYWKPPDEBGAQCPPICTHSCVDLDKXGCPAQORASPLTISAVVG 660

QY 660 PSQVKPDLSTMPYWKPPDEBGAQCPPICTHSCVDLDKXGCPAQORASPLTISAVVG 660
 DB 660 PSQVKPDLSTMPYWKPPDEBGAQCPPICTHSCVDLDKXGCPAQORASPLTISAVVG 660

QY 720 RYKXVLSGSAFGYVYGIWIPDEENKIVAIKRENTSPANKKEIIIDEAVVAGVSP 779
 DB 720 RYKXVLSGSAFGYVYGIWIPDEENKIVAIKRENTSPANKKEIIIDEAVVAGVSP 779

QY 781 YVSRLLGICLTSTVQVLTQMPYGCCLDHRHRRGLSGQDILMNCVQIAKMSLTLEVR 839
 DB 781 YVSRLLGICLTSTVQVLTQMPYGCCLDHRHRRGLSGQDILMNCVQIAKMSLTLEVR 839

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMMALESILRRFT 900

QY 900 HOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLEKGRLEPOPICTIDYVMIVKMM 960
 DB 900 HOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLEKGRLEPOPICTIDYVMIVKMM 960

QY 961 IDSECRPRFRELVSFSRMAADPQRFVVIQNEDLGPASLDSTFYRSLLEDMDMDLVA 1020
 DB 961 IDSECRPRFRELVSFSRMAADPQRFVVIQNEDLGPASLDSTFYRSLLEDMDMDLVA 1020

QY 1021 EBYLVPOGGFFCDDPAPAGAGVYHHHRSSSTRSGGDLTLGLFSEEBAPRSPAPSEG 1080
 DB 1021 EBYLVPOGGFFCDDPAPAGAGVYHHHRSSSTRSGGDLTLGLFSEEBAPRSPAPSEG 1080

DB 1020 EBYLVPOGGFFCDDPAPAGAGVYHHHRSSSTRSGGDLTLGLFSEEBAPRSPAPSEG 1080
 QY 1081 AGSDVFDGLMGAAKGLQSLPTHDPFLQRYSEDPYPLPSEFTDGYAPLTCSPQPEYV 1140
 DB 1080 AGSDVFDGLMGAAKGLQSLPTHDPFLQRYSEDPYPLPSEFTDGYAPLTCSPQPEYV 1140

QY 1141 NOPDYAPQPPSPREGPLPAPAPAGATLRR-----AKTSPGKGVYKDVFAFGAVENPE 1195
 DB 1140 NOPDYAPQPPSPREGPLPAPAPAGATLRR-----AKTSPGKGVYKDVFAFGAVENPE 1195

QY 1196 YLTPGGAAPQPPAPAPAPAFDNLVYWDQDPERKAPSPSTFGTPTAENPEYGLDVPV 1255
 DB 1200 YLAPRGAPAPQPPAPAPAFDNLVYWDQDPERKAPSPSTFGTPTAENPEYGLDVPV 1255

RESULT 2
 Q9QX70 PRELIMINARY; PRT; 1209 AA.
 AC Q9QX70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=9025868; PubMed=2342466;
 RA Patch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue."
 RT Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Patch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Earg H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: M37394; AAF14008.1; -.
 DR HSSP: P11362; 1RGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR01245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00669; pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR PRODOM: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 KM APP-binding; Receptor; Transferrase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 56FE78EC1B773 CRC64;

Query Match 45.8%; Score 3120; DB 11; Length 1209;
 Best Local Similarity 49.6%; Pred. No. 4-4e-228;
 Matches 634; Conservative 167; Mismatches 362; Indels 114; Gaps 26;

QY 3 LALACRWGLLALLPGAASTQVCTGDMKLRJSPASPTHLDMLRHL YGCGQVVGNTL 61
 DB 3 LALACRWGLLALLPGAASTQVCTGDMKLRJSPASPTHLDMLRHL YGCGQVVGNTL 61

```

Db      15 LALCAG-----GALEEKVCQGTSTNRLTQGTGFEDHFLSLQRMFNNCVLLGNLE 66
Qy      62 LTYLPTNASTLFLQDIOEVOQYIKANSKFIGITELORLIRVGTQLPEDNALAVLNGD 121
      67 IYVQQRNDLSFLKTIQEVAGYVLLALMTVERIPIENQIIRGNALVENTYALAVLEN-- 124
Qy      122 PLANTTPVTGASPGGLRELOLSLTELKGGVLLQORNPOLCYQDTILMKDI FHKNNOLAL 181
      125 -----YGNKTKLRELPMNRLQELIGAVRFSNNPILCNMTETIOMRDIV--ODVELSN 175
Qy      182 TLIDTNRS-RACHPSPMCKSGRCWSESSEDCOS:TRTVACAGGA-RCKGPLPTDCCHQ 239
      176 MSMDYGRHLTGCPKCDPSCPNSSCWGRGENCQKTKICAQCSRRCRGRSPSDCCHNQ 235
Qy      240 CAAGCTGPKHSDCLALPHNHSIGELHCPALVTYNTIDTFESMNPGRYTFGASCYTAC 239
      236 CAAGCTGPRESDCLVCHRFDEATCKDTCPPLMYNPTTYQMDVNPGEKYSFGATCYKCC 235
Qy      300 PNYVLTSDVGSCTLVCPILHNOEVTAEQTCRCCKSCRCARVCYGLMELREVAATSA 359
      296 PNNYVTTDHGSCVRACGPDYIEV-EDGVSKCKKCDGCRKVCNGIGIGEFKDTLSINAT 354
Qy      360 NIOEPAGCKKIFGSLAFLPESEFDPASNTAPLOPEOLQVETLEETIGYLYISAMPDL 419
      355 NIKHFKYCTAISGDHLIPVAFKSDSFRTPPLDBRELIKTVEKITGELLQAMPENW 414
Qy      420 PDLVFPQNLQVIRGILINNGAYSLTLOGLGISWIGLSLRELSGLALIHNNHLCFVAT 479
      415 TDLHAFENLEIRGTHKQGFSLAVWLINTSLGLRLKESIGDVIIGNRLCYANT 474
Qy      480 VPMQDLFRNPHQALHTANREDECVSGELACHQICARGHWMG3PPTOCVNCSCQFJGQ 539
      475 IMKKLFGTPNCKTKIMNNAEKDOKANHNHNCPLCSGSGWGEPEPDCVSCQVNSGRK 534
Qy      540 CYEBCRVUQGLPREYVNAHCLPCHPECQPNQSVTCFEPADOCVACAHYKCPFCVAR 539
      535 CVDKCNILLEGEBREVENSECIOCHPECLPQTNMITCTGSRPDMCICAHYVQPHCVKT 534
Qy      600 CPSSGVKPDLSYMPYKFPDEGACQPCGINCTHSCVDLDXKCAEQARAP-LTSTVSAV 638
      595 CPSSGIMGENNTL-VKFPDANNVCHLCHANCTYCAGBGLKGC--QQPEBKPSTATGI 651
Qy      659 VGILLVVLGVVFGI-LIKRQOKIKRYTMRLLOETELVEPLTPSGAMPNOQMRLIKE 717
      652 VGLLFLIVV-VALGIGLFMRRLQVLRKKTARLLQERELVEPLTPSEANQHLRIKE 710
Qy      718 TELRKVXVLGSGAECTYVKGWIPDGENVKIPVAKVLRENTSPXANKELIDBAYVAGV 777
      711 TELFKIKVLGSGAECTYVKGWIMEGEVVKIPVAKELRENTSPXANKELIDBAYVAVSV 770
Qy      778 GSPYVSRLLGICLNTSTVOLVTLMPYGLLDHVENGRIGSODLLMWCQIKKGSYLE 837
      771 DNPVCRLLGICLNTSTVOLVTLMPYGLLDYVAEHNDNGSVLLMWCQIKKGNAYLE 830
Qy      838 DVLVHRDLAARNVLVSPNVTCTDGLARLIDITEYHADGKVPYIKMALESILR 897
      831 DRLVHRDLAARNVLVKTPOHVKTIDFGLALGAEEKXVHAEGKVPYIKMALESILR 890
Qy      898 RFTQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPOPICTIDVWIMTK 957
      891 IYHQSDVWSYGVTVWELMTGSRPYGIRASELSLLEKGERLPOPICTIDVWIMTK 950
Qy      958 CMWIDSECRPFRELVSEFSMAADPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDMGD 1016
      951 CMWIDASRPFRELVSEFSKADPQRYLVIQDERMHLPSPTDSFYRSLMEEDMGD 1010
Qy      1017 LVDAEYLVPOQGFPCDPAPAGAGWVHHNRSSSTSGGDLTLGLEPSEEAAPSPLA 1076
      1011 VVDAEYLVPOQGF-----NSBST-----SRPSTL 1036
Qy      1077 PSAGASDPDGLGMAKAGLQSLPTHDSPLQRYSEDPVLPESET--DGVAPLTC 1134
      1037 SSLANSN-----SSTVACINRNGSCRVKEDAFQRYGSDPTSVLTEDNDIDFTL----- 1086

```

```

Qy      1135 POPEYVQDVPPOPPSPREGPLPAPAPAGATLEBAKTLSPGKNQVWDVPAFGAVENP 1194
      1087 PVPEYINQ-SVPRKAGSVQNPVHNGDLH-----AGRDHLION--PHSNASNP 1135
Qy      1195 EYL-TPQGAAPQPPHPPAFSPAFDNLTYWDQ-----DP-----PERGAPSTF 1237
      1136 EYLNTAQ-----PTCLSGSFGSALMIQKSHQMSLDNPDYQDFFPKRAKPNCIF 1186
Qy      1238 KGTPTAENPEYGLDVP 1254
      1187 KG-PTAENAEYLRVAP 1202
Db

```

```

RESULT 3
Q9EP98 PRELIMINARY, PRT, 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVATC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampiland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maille N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schell C., Pearshall R.S., Green P.J., Yee D., Lampiland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maille N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF275366; AAC28045.1; -.
DR EMBL: AF275365; AAC28045.1; JOINED.
DR EMBL: AF275367; AAC24386.1; -.
DR HSSP: P11362; IFGK.
DR MGD: MGI:95294; Egfr.
DR InterPro: IPR000345; CytC heme bind.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR007119; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR PRINTS: PR00109; TRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SMO0261; Fu; 5.
DR SMART: SMO0220; S_TKc; 1.
DR SMART: SMO0219; TyrKc; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

```



```

QY      161 LCYQDTLLMKDIFHKNNQALTLIDTNSRACHPCSPMKCKSRGCESSDDCQSLTRTVVC 220
DB      3 LCPADTIHMODIRNPMWASNFLLVPTNCGGCGCKSCCTG-RCMGPTRMHCQTLTKTVC 61
QY      221 AGGC-ARCKGPLPTDCCHQCAAGCTGPRGSDCLACIHNHSGITELHCPALVTYNTDTF 279
DB      62 AECDCGRCTGFPYVSDCCHRECAAGCGEPDTCFCAMFNDSGACVTCQCPOTFPVYNPTTF 121
QY      280 ESNPNEGRTYFGASCVTACPNYLTSTDVGCTLVCPHNOEVTAEQTCCEKSKPCA 339
DB      122 QLEHNHNAKYTAGAFCKKCPHNHF-VDSGCVRAQSPSKHEV-BENIKKCKCTDTC 179
QY      340 RVCYGLGMEHLREVRVTSANIQEPAGCKITPGSLAFPESDGDPASNTPLPDEQLQV 399
DB      180 KACDGIQTSLSVAQVDSNDIKFNCCKINGNLIPLVTGIHGDPRYHTIAINPEKINI 239
QY      400 FETLEETIGLYISAMPDLPDLVSFQNLQVIRGRILHNAGASLTLOGLSWGLSLR 459
DB      240 FQTVREITGYLINIGSPENMTDFRVFSNLVTIGBALYSGLSLILKQGITSLQFQSLK 299
QY      460 ELGSGIALIHNHTLFCVHTVPWQLEFNNHQALHTANPEDECVSEGLACHQLCARGH 519
DB      300 QISAGNIYITDNSNICYHTVMTSLFSTPSQKVIHNKKAENCTADGMVCNELCSSDG 359
QY      520 CMGPRPTOCVNSQPLRGOECVCECRVILQGLPREVYNAHCLPCHPEQCP-QNSVYTCFG 578
DB      360 CMGPRPDQCLSKRPIRGKTCIESCNLVDEFRFPANSVCMECDPQCEKXEDMTTCYG 439
QY      579 PEADOCVACAHYKDPFCVAPCSPGVKBDLSYMPIWKEPDEGACOPCPINCTHSCVDLD 638
DB      420 PGPDHCTKCFHFKDPNCEKCPDGLQGANSP--IFKYADRECHPCHPNCTGCGCPRA 477
QY      639 DKGC-----PABGRASPLTISYSAVY-GILVVYVGVFGILIKRQCKIRKYT 666
DB      478 SHDCIYPTWTRQSTLPOHAR-TPL--IAAGVIGGFIIVIGLIFAYVYKRKS-K-KRA 533
QY      687 MRLLQETELVEPLTPSGAMPVQAOMRILKETELRYKVLGSAFGTVYGIWIPDGENV 746
DB      534 LRRL-ETELVEPLTPSGTAPQAQRLIKETELRYKVLGSAFGTVYGIWIPDEETV 592
QY      747 KIPVALIKVIRENTSKANKELIDEVYVMAVGSPYVSRBLGICLTSTVOLVTQIMPYGL 806
DB      593 KIPVALIKILNETTGKAVEMDEKLIMASNDHPLVRLGLVCSPLTQVLQMLPHGCL 652
QY      807 LDHVENRGRILGSDPLIMCMQIAKMSYLEDVLYHRDLAARVYLKSNHYKTDFTGL 866
DB      653 LDYVHEHKNIGSOLLIMVQIAKMYLEERLYHRDLAARVYLKSNHYKTDFTGL 712
QY      867 ARLLDIDETEYHADGKVPFKMALESILRRFTHQSDVMSYGVTVMLMTFGAKPYDGI 926
DB      713 ARLEGEDEKEYNADGKVPFKMALECIHYRKFHQSDVMSYGVTVMLMTFGKPYDGI 772
QY      927 PAREIPDLLEKGERLPOPICTIVYMIWYKCMWIDSECPREPLVSEFSRAARDPQR 986
DB      773 PIREIPDLLEKGERLPOPICTIVYMIWYKCMWIDSECPREPLVSEFSRAARDPQR 932
QY      987 VVIQNEQ-LGPASPLDSTFYRSLLEDDMDGLVDAEYVLPQOGFFCPDPAPAGGMVNH 1045
DB      833 LVIOGDDRMKLPSPENDSKFQNLDEEDLEMDMAERYLP-QAFNIPPIYTSRRIRDS 891
QY      1046 RHRSSSTPSGGGDLTLGLEPSEEARPS--PLAP-SGASADVFDGDLGMAKGIQSLP 1102
DB      892 NRNOFTVRDGGYAAEQGV-PMYPYAPGCIIPFAVAAGATAEIFEDTCNGTLRKQVATL 950
QY      1103 THDSPLOQVSEDEPTVLPs-----ETDGYVAFLTCSQPEYVQAPVAPQPPSPRSG 1155
DB      951 AKEDSSTQRYADPTVFLPERVIRGELDEDOGYMTPMRDKXTIYLAIVENPVSRRKRG 1010
QY      1156 PLPAA-RPAGATLERAKTILSPGKGVYKVF-----AFGAIVENPEYLTPOGGAAPQ 1206
DB      1011 DLQAVNDPEYHN-----APNGQPKADEYVNEPLVLTNTFANTLENAYL-----K 1055
QY      1207 FHPFAPSPAFDNLVYWDOPPERGA--PSTFKGPT-----AENPEYL 1249

```

```

DB      1056 NNLPKAKKAFDPNDPMNHSLEPRSTLGHFDVLQESTKYFYKONGRIPIVAENPEYL 1114

RESULT 5
Q9YH40 PRELIMINARY; PRT: 1165 AA.
ID Q9YH40
AC Q9YH40;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorphi;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Minkler C., Wellbrock C., Gomez A., Duesch J.,
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RL overexpression and mutational alterations.",
RN Oncogene 16:1681-1690(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Scharlt M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53471; AD010500.2; -.
DR HSSP: P11362; IPGK.
DR InterPro: IPR000145; Cyrc_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D8771A74E CAC64;

Query Match 39.7%; Score 2703; DB 13; Length 1165;
Beet Local Similarity 45.3%; Pred. No. 2,1e-196;
Matches 576; Conservative 163; Mismatches 393; Indels 140; Gaps 29;

QY      1 MELAALCRWGILLALALPPG-AAST---QVCTGDMKRLPAPFERHLMRLHYGGCOV 55
DB      4 LELEL-----LILLLSIGRCCTDPKRVCGTSNQMTN--LDNHYYLKKMKMYSGCNV 56
QY      56 VQGNLELTYPTNASUSFLDIDQEVQOYIKANSKFIGITELQRLIVRGTOLPEDNYALA 115
DB      57 VLENLEITYQENQDLSFQISQIEGVGYVLIANNVSTIPLVVLRIIRQNLVEGNGFTLL 116
QY      116 VLDNGDPLNNTTYTASPGGLREIQRLSTELTKKGVLIQRNPQLCYQDTLLMKDIFHK 175
DB      117 VMSYVQK-NESSP--DYQVGLKQLQSLNLTTELISGSGVAVSNPCLCNVETINMWDIVDK 173
QY      176 NNQALTLIDTNSRACHPCSPMKCKSRGCESSDDCQSLTRTVACAGC-ARCKGPLPTD 234

```

```

Db      174 TSNPTMNLIPHAFERCCQCKDCPCVNGSCWAPRGHCQKFTLLLAECQCNRCRCPKID 233
Qy      235 CCHQCAAGCTGPKHSDCLAFHNHSGICELHCPALVTYNTDTFESMPNEGRTFGAS 294
Db      234 CCHHCAGGCTGPRATDCLACRDNFDGTCXKDTCPPKIYDVSQVVDNPNIKYTFCA 293
Qy      295 CTTACCPYNTLSTDVSGCTLVCPHNOEYTAEDGTQRCCKSCSPCARVCYCGMEHLRYR 354
Db      294 CVKECPSNVYVE-GACVRSAGAKMEVD-ENGKSKCKPCDGVCPKVDGIGISLNTI 351
Qy      355 AVTSANIIOEFAGCKKFFGSLAPLESFDDPASNTPAQEQLQVFETLEBITGYLYSA 414
Db      352 AVNSTNGSFNSNCTKINGDILNRSFEEDPHYKIGPMDPEHMLMTLVKELTGLVIMW 411
Qy      415 WPDSLPDLVSFQQLQYIRGRIILHNKYS-LTQIGISWLSGRSLREISGLALHNHTH 473
Db      412 WPMNTSLSVFQNLLETIRGRTTFSRGSFVVOVSHLQWLGSLSKREVSANVILKNTPO 471
Qy      474 LCFVHTVPMQDLFRNDHQAHLHTANPEDECVGEGLAGHOLGARGHCWGPPTOCVNSQ 533
Db      472 LRYASTINMRRLFRSDQSEVDART-----ENQTCNNESESDCGPPTMVCGLH 524
Qy      534 FLRGQCEVEECRYLQGLPREYVNAHCLPCHPECCPQNGSVTCFPEPADQVCAHYKDP 593
Db      525 VDRGRCVASCNLLQGEPRBAQYDGRQVQCHOECVQTDLSLTCYGPGRANSCKAHFQDG 584
Qy      594 PFCVACPSGVKPDLSYMPIMKFPDEEGACOPCINCHSCVDLDDKCPAEGASPLTS 653
Db      585 PCICPRCPHGMADGDTL-IMKADMGCCQCHONCTQCGSGPLSCGRD-IVSHSL 642
Qy      654 IYSAVVGILLVVLGVVFSILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAMR 713
Db      643 AVGLVSGLLITVALLIIVLLRRRIK-RKRTIRLLQEKELVEPLTPSGAPQAOAFLR 701
Qy      714 ILKETELRKVYLGSGAFGVYKGIIPGENVKI.PVAILKYLRENTSKAKELIDEXYV 773
Db      702 ILKETELFKDRVLSGAFGVYKGLNPPGENIRI.PVAILKYLRENTSKVQOEVLDEXYV 761
Qy      774 MAGVSPVYSRLIGLICLTSTVOLVTQLMFYGCLLDHVENRGLSODILMCMQIAKMG 833
Db      762 MASVDHPHYCRLLIGLICLTSAVOLVTQLMFYGCLLDYVHOHQRICGQMLMWCQIAKMG 821
Qy      834 SYIEDRLVYHRDLAANLVKSPNHYKIDPGLALLDIDETEHYADGKYPIMMALES 893
Db      822 NYLEERHLVHRDLAANLVKSPNHYKIDPGLSKLLADEKEXYQAHGKXVPIKMMALES 881
Qy      894 ILRRFTHOSDVMSYGVYVWEIMLTGAKFYDGI.PAREIPDLLEKGERLPPOPICTIDYV 953
Db      882 ILQMTYTHOSDVMSYGVYVWEIMLTGSKRYDGI.PAKETASVLENERLPOPICTIEYV 941
Qy      954 IMYKCMWIDSECPRRRELIVSEFSKMAPPOPFVVIQNEIDGAPASPLDSTYRSLLEDD 1013
Db      942 IILKCMWIDPSSPRRELIVGEFSQMAPPSRYLVIOG---NLPPSRRR.FSRLSSDD 998
Qy      1014 MGBLVAAEEYLVQCGFPDPAPAGAGVHHRHSSSTRSGGDLUTGLPSEBEAPRS 1073
Db      999 ---DIVDADEYLL-----RYKRIK-RQS-----E 1019
Qy      1074 PLAPSEGASDVFDGDLGMAKAGLQSLPHTDPSLPQRYSEDPIV-PLPSTDGVAPLT 1132
Db      1020 PCIPPNH-----FVRENSIALRYISDPTQNALAKDLGGH----- 1054
Qy      1133 CSFQPEYVNOPIVRPQ-----PSPRE-----GPLP-AARPAATILERAKTISPGNG 1179
Db      1055 -----EYVNOFSGSETSRLSDIYNRYEDULTGKWPVLSISGEATNFSREYVNTQNS 1109
Qy      1180 VYKDVAFAGAVENPEYLTPOGGAAPQHPHPAPAFADNLYYMDQDPERGAPSTFGK 1239
Db      1110 L---PLVSSGSMDDPY---QAG-----YQAAF-----LPQGALTGKGMF 1144
Qy      1240 TPTAENPEYIGL 1251

```

```

Db      1145 LPAENLEYIGL 1156
RESULT 6
ID P79754 PRELIMINARY; PRT: 1326 AA.
AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ekb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=11033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wrt1 locus of Fugu
RT rubripes.";
RL Genome Res. 9:251-258 (1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fy_3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ SEQUENCE 1326 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match      33.4%; Score 2275; DB 13; Length 1328;
Best Local Similarity 40.2%; Pred. No. 9,1e-164;
Matches 517; Conservative 152; Mismatches 427; Indels 190; Gaps 31;

```

```

QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLELQSGALITHNTHLCFYHTV 480
DB 404 DLSVFSSLTTLICRSILFKRFSILWVRIFLTLSTGLRSLEISDSGVISQAHLCYHNTV 463
QY 481 PMDQLFRNH-QALLHTARPEDECGEGLACHQICANGHCKGPPPTCCVNCQSLRGQE 539
DB 464 NMTQLFRGRVRANSINSRPMACVADRVCDPLCSGSGCWGPDPDCLSCRYSRHGT 523
QY 540 CVEECRVLOGLPREYVNAH-CLPCHPECOPNGSVTCFGEPAADQVACAHYKDPFECVA 598
DB 524 CVAGCHFNNGIIPREFAGLNGVACHPEKPEPTGASCTGPADEBMACTFRGPRICMS 583
QY 599 RCPSSVYKPLSTWPTWKFPDEBGAQCPCTINCHSCVLDKGCPEAGRASPLTSISAV 658
DB 584 SCPAGVN-DGEGKLIKFPNREGHCEPHONCTQCGSPGINDC---LEAARLTSSGQI 639
QY 659 VGILLVVLGVVF-----GILIKRQOKIRKYMRLLOTELEVEITLTSAGMPNOAQ 711
DB 640 TGIALGVPAGLIFCLVLPFLGMLYRGLAIRKRAMRYLESSESFEPLGF-GEKGTKVH 698
QY 712 MRLKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTPSKANKELIDEA 771
DB 699 ARLKPSDLRKIKPLGSGVFTVSKGFWLPEGETYKIPVAIKTIDSSGRQTFTETDHL 758
QY 772 YVMAGVSPYVRILGICLTSTVQVLTQLMYGCILDEHVRNRLGSLQDLNMCQIAK 831
DB 759 LSMGSLDHPYIVLLGICPGTCLQVLTSSHGLLEIRHKHSLDPORLNMCQIAK 818
QY 832 GMSYLEDELVLRDLAARNVLVKSFNHYKIDFGIARLLDIDETEHADGKVPKIMMAL 891
DB 819 GMYLEBRHVHKNLAARNILILKNDYQVQISDYADLLPYDDKKYVSEKRTIKMAL 878
QY 892 ESILARRFTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIDLEKEBRLEPPPTCTIDV 951
DB 879 ESILARRFTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIDLEKEBRLEPPPTCTIDV 938
QY 952 YWIMVYKCMWIDSECPREPRELSEFSRMAPRPFVYQNDLCPAPBLDSTFRSLLED 1011
DB 939 YWIMVYKCMWIDSECPREPRELSEFSRMAPRPFVYQNDLCPAPBLDSTFRSLLED 982
QY 1012 DDMGDLVAEEYLVPOQGFPCPDPAAGVHHRRSSSTRSGGDLTLGLEPSEEAR 1071
DB 983 SCMGFFL-----RRGSR---GLLEADLEDEDE-- 1008
QY 1072 RSLAPSEGAGSDVFDGDLGNG--AAKGLQSLPTHPSPIQ-----RYSEDPVY 1118
DB 1009 -----GLGDRFATPSLDQSPSWSTSPQINSYVMTQLRYD----- 1044
QY 1119 PLPSETDGYVAPLTGSPQ-EPVNO-----PDVPPQSPPREGPI--PAAR 1161
DB 1045 -FAVSGGHIGYLPSPSPVDITIRQLWYQSRSLSSVRLPDRSAFRSSSREALCEDGAQ 1103
QY 1162 PAGATLEBAKTLSPGKNVAVDVAFGAVENPEYLLPQCGAAPQHPHPPAPSPADNLY 1221
DB 1104 CAGIFRVV-----FSSERGN-----PQGG----- 1122
QY 1222 YWDQDPPERGAPSTFKGTPTAENPE 1247
DB 1123 --QQRKLTASSPSPSKTWADEDE 1146

```

RESULT 7

```

Q9BIH9 PRELIMINARY; PR7; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

```

OC Anopheles.
OX NCBI_TaxId=7165;
RN 1
SEQUENCE FROM N.A.
RC STRAIN=GA.
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AJ301655; CAC35008.1; ..
HSSP: P11362; 1FGK.
DR InterPro: IPR000345; CysC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002280; Set-ctr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 7.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 29.5%; Score 2011.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 11e-143;
Matches 467; Conservative 197; Mismatches 388; Indels 385; Gaps 36;

```

```

QY 549 GLPREY-VNARHCLPHPECOQONGSVTCFGEADQVCAHYKQPPCVARCP----- 601
DB 507 SLRLYVSVDKSCGDCQECQCKD-----FCYGRPNQCSQMMVYKGRFCVACPTTKHAM 561
QY 602 -----SGVFPDLSTYMPIWKFPD----- 618
DB 562 NGTCINCHKTVCGRGPRDTIAPDGCISCDKAIIGSDALIERCLMKKSDCPDGYSDVYL 621
QY 619 -EEG----- 621
DB 622 QEEGFLKQLSGKAVCRKCHPKCKCTGYGFHDFQCECTGYKKGEQCEDECPQFYANEE 681
QY 622 --ACQPCPINCT-----HSCVDL-----DD-----XGCPAEQ----- 646
DB 682 TRICLPHQECRCHGLGDDHNECRNLKFBEGDPYDNTFTTCVANCASHYKKEFQBEA 741
QY 647 -----RASPLTISVAVVGIILVVVLGVVFI---LKRQOKIRKXTM 687
DB 742 GKIGPYCSADSMOSGRIEFPOTGVKIVMSVALLILCVFGIAFVLESRHKKKDAVKM 801
QY 688 RRLLOETELVEPLTPGAMPNQOMRILKETELRKVYLSGAFGVYKGIKIPGCEVYK 747
DB 802 TMAAGCESEPLRPSPVGNPLTKRIIEABIRKGVLSGAFGRVFRKGVAMPGESVYK 861
QY 748 IYVAIKVIRENTSPKANKELIDEAYVAVGSPYVSRLLIGLSTVQLVQLMPEYGLL 807
DB 862 IYVAIKVIRENTSPKANKELIDEAYVAVGSPYVSRLLIGLSTVQLVQLMPEYGLL 921
QY 808 DHYRENRGRIGSODLNNMCQIAKMSYLEDVRLVHRDLAANVLKSPNRYKIIDFELA 867
DB 922 DYVRNNKDKIGSALLNWSQIARGMAYIEERLVHRDLAANVLVQPSCKIVFVFLA 981
QY 868 RLIDIDETEHADGKVPKIMWALLESILRRRFTHOSDVSYGVTWELMTFGAKPYDIP 927
DB 982 KLIDFDSDEYRAGGMPIKIMWALLESILRRRFTHOSDVSYGVTWELMTFGAKPYDIP 1041
QY 928 AREPDLLEKEXELPQPICTIDVYIMWKCMIDSECPREELVSESRMARPOQFV 987
DB 1042 AKDVPELIEIGHKLPOPDICSLDVYICILSCVWLADAPTRKQLAETFAEXARDPGRYL 1101
QY 988 VIQNEIDGPASPLDSTFYRSILEDDPMGLV----- 1018
DB 1102 MI-----PQDKMRRLPSTYNODEKDLIRTLARVAMAAAAAAGASNVDPSTIA 1152
QY 1019 DAEYLVPOQGFPCPPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEARPS----- 1073
DB 1153 ETDVYLQPKTRPSIMLPGBSA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP--SEGASDVFDDGLGWAAGKGLQSLPTHDPSPFLQRYSDPTVPLPSETDGYVA 1129
DB 1189 DPLKPDDETGHGKEV-----GVGGIR-----LNDPLDEDDYLM 1222
QY 1130 PLTCSPOFEVYNQPDVRCPPSPREGRPLAARPAATLERAKTLSPGKGVKQVFAFGG 1189
DB 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243
QY 1190 AVENPEYL-----TPQGAAPQHPAPAPAFNLDVLYWDDPFBRCAPSTFKGT 1240
DB 1244 SVDPPEYLMGSTALAGLAGAGSWG--PHTP-----PPTPNGM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

RESULT 8
Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Heistatin.

```

CN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLIB347E2D030C CRC64;

```

Query Match 26.4%; Score 1797.5; DB 4; Length 419;
Best local Similarity 86.3%; Pred. No. 3.3e-128;
Matches 340; Conservative 7; Mismatches 40; Indels 7; Gaps 2

```

QY 1 METALCRNGLLALLPFGAATGYCTGTDMLRLPASPETHLDMRLHYGQCYVQGNL 60
DB 1 METALCRNGLLALLPFGAATGYCTGTDMLRLPASPETHLDMRLHYGQCYVQGNL 60
QY 61 ELTYLPTNASTFLQDIDCEVOQYKANSKFIGITELQRIYRGQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASTFLQDIDCEVOQYKANSKFIGITELQRIYRGQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYASPGGLRELQRLSLTEILKGVLLIQRNPOLCYQDTILMKDIFKNNOLA 180
DB 121 DPLNNTPTVYASPGGLRELQRLSLTEILKGVLLIQRNPOLCYQDTILMKDIFKNNOLA 180
QY 181 LTLIDNBRACHPCSPMKSGSRGSESSDQSLTRVCAAGCARCGPLPTDCHEQC 240
DB 181 LTLIDNBRACHPCSPMKSGSRGSESSDQSLTRVCAAGCARCGPLPTDCHEQC 240
QY 241 AAGCTGPKASDCLACLFHNSGICELHPALVTYNTDFTESKPNBGRYTTGASCVTACP 300
DB 241 AAGCTGPKASDCLACLFHNSGICELHPALVTYNTDFTESKPNBGRYTTGASCVTACP 300
QY 301 YNLTSDVGSCTLVGLNHQEVTAEDGQRCCKSKPCARVYGLGMEHLAEVAVTSAN 360
DB 301 YNLTSDVGSCTLVGLNHQEVTAEDGQRCCKSKPCARVYGLGMEHLAEVAVTSAN 360
QY 361 IOEFAGCKKIFSLAFLPESFGDASNTAPLPQ 394
DB 356 LRMQPG--PAHPLVLSLRPSMDLVARFSLPLAP 387

```

RESULT 9
Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 7, 5e-124;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

```

QY      889  MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLKEXGERLPQPICT 948
DB      1  MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLKEXGERLPQPICT 60
QY      949  IDVYMIWVKCMIMDSGCRPRFRELVESESRMARDPQRFVVIQNEIDLGPSPIDSTFYASL 1008
DB      61  IDVYMIWVKCMIMDSGCRPRFRELVESESRMARDPQRFVVIQNEIDLGPSPIDSTFYASL 120
QY      1009 LEDDDMGDLVDAAEEYLVPOGFCPCDPAPGAGGMVHRSSSTSGGQDLTLGLIEPSEE 1068
DB      121 LEDDDMGDLVDAAEEYLVPOGFCPCDPAPGAGGMVHRSSSTSGGQDLTLGLIEPSEE 180
QY      1069 EAPRSPLAPSEGAGSDVFDGLMGAKGLQSLPTHDPSPLOGYSEDPTVPLPSETDGYV 1128
DB      181 EAPRSPLAPSEGAGSDVFDGLMGAKGLQSLPTHDPSPLOGYSEDPTVPLPSETDGYV 240
QY      1129 AETCSPOPEYVNOQPVRRPQPPRREGPLPARPAGATLERAKTISFGKNGVYKDVAFAG 1168
DB      241 APLACSPQPEYVNOQPVRRPQPPRREGPLPARPAGATLERAKTISFGKNGVYKDVAFAG 300
QY      1189 GAVENEPEYLTPOGGAPOPFPAPFSPAFDNLVYMDQPERGAPPTSTFKGPTAENEY 1248
DB      301 GAVENEPEYLTPOGGAPOPFPAPFSPAFDNLVYMDQPERGAPPTSTFKGPTAENEY 360
QY      1249 LGLDVVPV 1255
DB      361 LGLDVVPV 367

```

RESULT 10
 086712 PRELIMINARY; PRT; 729 AA.

AC 086712;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Polyprotein.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSSP; P03322; IAS6.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02813; Retro_M; 1.
 DR Prodom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFD63 CRC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 5, 9e-122;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15

```

QY      569  PONGSVTCFGEADQCVACAHYKDPFCVAPRCPSPGPDLSYMPIKMPDEGACPCPI 628
DB      141 PEETATPRTGP--DHCMKCAHFIDGPHCVKACBAGVLGENDTL-VWKYADANAACQLGHP 197
QY      629  NCHSCVDDLDKCCPAEQAASPLTISVSAV-V-GILLVVLGVVFGILLRPOOKTRKYM 687
DB      198 NCTRGKGPGEGCP--NGSKTPSIAAGVAGGLCLVVGIGIGLYLRR-HIVRKRL 253
QY      688  RLLOETELVEPLTPSGAMPNOAMRLKETELRKVKVLSGAFGVYKGIWIPGENVK 747
DB      254 RLLORELEVEPLTPSGEAPNOAHLRLKETEFKVKVLSGAFGVYKGIWIPGEEKK 313
QY      748  IPAIKVLRNLTSPKANKELDEAYVMAVGSPPYRLLIGLITSTVQLTQMLPFGCLL 807
DB      314 IPAIKVLRNLTSPKANKELDEAYVMAVGSPPYRLLIGLITSTVQLTQMLPFGCLL 373
QY      808  DHVRENRGRLSQDLINMCMQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKTDFGLA 867
DB      374 DYIREKDNIGSVYLLNMCVQIAKGMNLEERLVHRDLAARNVLKTPQHVKITDFGLA 433
QY      868  RLIDIDETEVHADGKVPILKMALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGI 927
DB      434 KILGADEKEVHAEGKVPILKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI 493
QY      928  ARILPDLKEXGERLPQPICTIDVYMIWVKCMIMDSGCRPRFRELVESESRMARDPQRFV 987
DB      494 ASEISVLEKGERLPQPICTIDVYMIWVKCMIMADSPPKRELIASFVKARDPPEYL 553
QY      988  VIQ-NEDLGAPSLDSTFYRSLLEDQDMCDLVDAAEEYLVPOGFCPCDPAPGAGGMVHR 1041
DB      554 VIQDERMHLPSPTDSKFRYRLMEEDMEDIVDABEYLVPRQGF----- 598
QY      1047 HRSSSTRSGGDLTLGLEPSEEARPSL-----APSEGAGSDVFDGLMGAKGLQSL 1101
DB      599 -NSPST-----SRTPLSLSLSTSNNSATNCID-----RNGGCH 631
QY      1102 PTHDPSPLOGYSEDPTVPLPSET--DGVVAPITCSPOPEYVNOQPVRRPQPSRREGPLPA 1155
DB      632 PVREDSFVRYSDPTGNLLESDIGFL-----PAPEYVNO-LMPKKKS----- 675
QY      1160 ARPAGATLERAKTISFGKNGVYKDV-----AFGGAVENEPEYL 1197
DB      676 -----TAWVQNVQIVNNISLTAISKLPMDSRYQNSHSTAIVDNPXYL 715

```

RESULT 11
 086714 PRELIMINARY; PRT; 567 AA.

AC 086714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE v-erbB protein (Fragment).
 GN v-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; -
 DR HSSP; P13362; IAFK.
 DR InterPro; IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase: 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 KW NON TER
 FT
 SQ SEQUENCE 567 AA; 63390 MW; CSD9CBAVAD725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 5.7e-122;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GREADQVCAAHYKDPKPCVACRCGKPKDLSYPIKPFDEBACQCPICNTGSCVDL 637
 DB 1 GP--DHCKCAHFIDGPHCVACAFAGLGENDTL-VKYNADANAVQQLCHPCTRCCKGP 57
 QY 638 DDKGPAPQASPLTSIVSAVY-GILLVVLGVVFGILIRKROOKIRKYMRLDQETEL 696
 DB 58 GLEGCP--NGSKPTSIAGVVGSLCLVVGIGLGLYLR--HVRKRTLRLLQEREL 113
 QY 697 VEPLTPSGAMPNOMRLKTELKRYKVYGSAGFTVYKGIWIPGENVKIPVALKVLK 756
 DB 114 VEPLTPSGEAPNOMRLKTELKTEFEKVKVYGSAGFTVYKGIWIPGEKVKIPVALKELR 173
 QY 757 ENTSPKANKELIDAEYVAVAGVSPYVSRLLGICLTSTVQLVQLMYPGCLDVRNRKR 816
 DB 174 EATSPKANKELIDAEYVAVASVDNPRVCRLLGICLTSTVQLVQLMYPGCLDVRNRKR 233
 QY 817 LGSODLLNMCQIAKNGSYLEDVRLVARDLAARNVLYKSPNHVKITDFGLARLLDDETE 876
 DB 234 IGSODLLNMCQIAKNGSYLEDVRLVARDLAARNVLYKSPNHVKITDFGLARLLDDETE 293
 QY 877 YHADGKVPKIMKMLESLIRRFTHQSDVSYGVYVWELMTFPAKYDGIIPAREIPDLLE 936
 DB 294 YHAGGKVPKIMKMLESLIRRFTHQSDVSYGVYVWELMTFPAKYDGIIPAREIPDLLE 353
 QY 937 KGERLPQPPICITIDVYIMVYKCMWIDSECRPFRELVESEFARMARDPQFVYIQ--NEDIG 995
 DB 354 KGERLPQPPICITIDVYIMVYKCMWIDSECRPFRELVESEFARMARDPQFVYIQ--NEDIG 413
 QY 996 PASPLDSTFYRSLLEDMDGLVDAAEYVLPQGFPCPPAPAGAGMVRHRRSSSTRSG 1055
 DB 414 LPSPTDSKFRYRTLEEDMEDIVDADEYLVPHQGF--NSPST-- 454
 QY 1056 GGDLLTGLEPSEERAPSP--APSEGAGSDVFDGDLGMGAAGKGLSLPTHDPSPQL 1110
 DB 455 -----SRPLLSLSATSNNSATNCID-----RNGQGHVYREDSEVQ 491
 QY 1111 RYSEDPVPLPSEF--DGYVAPLTCSPQPEYVQNPVRPQPSFREGPLPAARPAQATLE 1168
 DB 492 RYSSDPPTGNFLEESIDGFL-----PAPRYVQ--LMPKPS-- 526
 QY 1169 RAKTLSPKNGVAVDF-----AFGAVENPEYL 1197
 DB 527 -----TAVVQVQIYNNISLTAISKLPMDSRVQNSHSAVDNPEYL 566

RESULT 12
 ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN P3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OK NCBI_TaxID=9606;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.F., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RA "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318349; AAU55856.1; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00108; EF_HAND; UNKNOWN 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.3e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 889 MALESILRRRFTHQSDVSYGVYVWELMTFPAKYDGIIPAREIPDLLEKGERLPQPPIC 948
 DB 1 MALESILRRRFTHQSDVSYGVYVWELMTFPAKYDGIIPAREIPDLLEKGERLPQPPIC 60
 QY 949 IDVYIMVYKCMWIDSECRPFRELVESEFARMARDPQFVYIQNEDIGPASPLDSTFYRSL 100
 DB 61 IDVYIMVYKCMWIDSECRPFRELVESEFARMARDPQFVYIQNEDIGPASPLDSTFYRSL 120
 QY 1009 LEDDDMDGLVDAAEYVLPQGFPCPPAPAGAGMVRHRRSSSTRSGGDLTLGLEPSEE 106
 DB 121 LEDDDMDGLVDAAEYVLPQGFPCPPAPAGAGMVRHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPSPAPSEAGSDVFDGDLGMGAAGKGLSLPTHDPSPQLRYSEDPVPLPSETDGYV 112
 DB 181 EAPSPAPSEAGSDVFDGDLGMGAAGKGLSLPTHDPSPQLRYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSPOPEYVQNPVRPQPSFREGPLPAARPAQATLERAKTISPKNQGVKQVDFAFG 1181
 DB 241 APLTCSPOPEYVQNPVRPQPSFREGPLPAARPAQATLERAKTISPKNQGVKQVDFAFG 300
 QY 1189 GAVENPEYLPQGGAPQ-----HPPA--FSFAFDNL 1221
 DB 301 GAVENPEYLPQGGAPQ-----HPPA--FSFAFDNL 360
 QY 1221 YWVD--ODPPER-----GAPSTFKGTPTAEN 1245
 DB 361 WWTTCQCEPEGQVRRSPDVSSGREGLTSAQIKRMGPPRTTSRGTCARFN 410

RESULT 13
 ID Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Gag_v-erb-A_v-erb-B protein.
 GN GAG_v-erb-A_v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retrod viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90206603; PubMed=1969616;
 RA Bruskino A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 transforming potential of the oncogene v-erb-B";

Query Match	24.3%	Score 1653.5;	DB 15;	Length 962;
Best Local Similarity	51.7%	Pred. No. 1e-116;		
Matches 358;	Conservative 73;	Mismatches 142;	Indels 119;	Gaps 18

QY	1107	SPFGKRSDEPTPTPLBSMTDGYAALPLTCSQPEYVNOVDVRCPPSPREBCLPAAPACAT	116
Db	899	-----PVEDGFL-----PAPYVNO--LMPKKSTAMVONQIYNYISLT	936
QY	1167	-LEBAKTLSPCKRNGVVKDVFAGCAVENEYCL	1197
Db	937	AISKLPMSDRYON-----SHSTAVDNEPYL	961
RESULT 14			
ID	Q85468	PRELIMINARY;	PRT; 545 AA.
AC	Q85468;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)	
DE	Avian Erythroblastosis virus (T834) v-erbB gene.		
OS	Avian erythroblastosis virus.		
OC	Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.		
OX	NCBI_Taxid=11861;		
RN	[1]	_Taxid=11861;	
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=88217326; PubMed=2897102;		
RA	Scotling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;		
RT	"Common site of mutation in the erbB gene of avian erythroblastosis		
RT	virus mutants that are temperature sensitive for transformation.";		
RL	Oncogene Res. 1:265-278(1987).		
DR	EMBL; X06943; CA330024.1; -.		
DR	HSSP; P11362; 1PKG.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	Pfam; PF00069; pkinase; 1.		
DR	ProDom; PD000001; Euk_pkinase; 1.		
DR	SMART; SMO0219; TyrcK; 1.		
DR	PROSITE; PS00107; PROTEIN KINASE ATP. 1.		
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM. 1.		
DR	PROSITE; PS00109; PROTEIN KINASE_TYR. 1.		
KM	ATP-binding; Transferase; Tyrosine-Protein kinase.		
SQ	SEQUENCE 545 AA; 60899 MW; 140DCB8CCCA0F8AF4 CRC64;		
Query Match 24.2%; Score 1645; DB 15; Length 545;			
Best Local Similarity 54.9%; Pred. No. 1.9e-116;			
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;			
QY	578	GPEADQCAAHYDPPEFCVACRCPGKPLDLYMPIKFPDEEGACOPCPINCHSCVDL	637
Db	1	GP--DHCMKCAHFIDGPHCVACPAVLGENDTL-VKYNADANAVCOLCHPNCRCGCKP	57
QY	638	DDKCPAEGQASPLTSTVSAV-VGILLVVLGVVFGILLKRRQOKIRKYMTRLLQETEL	696
Db	58	GLECCP---NGSKTPSTIAAGVGGCLLVAGGIGIGLYLRR-HYVKRTLRLRLQEREL	113
QY	697	VEPLTPSGAMPNQAKNLIKETELRKVYLGSGAFVYGVGIWIPDENVKIPAIKVLRL	756
Db	114	VEPLTPSGEAPNQAHRIILKETEKVKYVLFQAFGVYGVGIWIPBEKXTIIPVATIELR	173
QY	757	ENNSPKNKEILDEAYYMAVGSPYSLIGICLTSTVOLVTOLMPYGCLLDHYRENGR	816
Db	174	EATSPKANKKILDEAYYMAVDNPHRCRLGICLTSTVQIITOLMPYGCLLDVIREHKN	233
QY	817	IGSODLLNMCQIAKGSYLEDVRLVHRDLAARNVLYKSPNHVKINDPGIARLLDDETE	876
Db	234	IGSQYLLNMCVQIAKGNVLYEEHGLVHRDLAARNVLYKTPQDYKIITDFGLAKQLGADKE	293
QY	877	YHADGKVPFKMMALLESILRRRTFHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIDLE	936
Db	294	YHAEKGKVPFKMMALSIILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLE	353
QY	937	KGSRLLQPPICITDVMYIMVYKCMNISQECRPPREIYVSEFSMAAPQCFVYIQ-NEDG	995
Db	354	KGSRLLQPPICITDVMYIMVYKCMNSADDERKFKRELIAPFSKQARDPEYLYVIGGERRN	413

```

QY 996 PASPLDSTFYRLSLDDMDGLVDAEYLVPQGFCCDPAPGAGMVRHRRSSSTRGC 1055
DB 414 LPSPTDSKFRITLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLLTGLSESEEAAPRSP-----APSEGAGSVDFDGLMGAKGLQSLPTDPSPLO 1110
DB 455 -----SRTPLLSLSLATSNNKNTNIDNNG-----H----- 481
QY 1111 RYSEDPYVLPSETDGYVAPLTCSPQPEYVNPQDPVPPSPREGPLPAAPACAT-LE 1169
DB 482 -----PYREDGFL-----PAPEYVNG--LMPKPKSTAVYQIYVYISLTATISK 523
QY 1170 AKTLSPGKGVKDVAFAPGAVENPEYL 1197
DB 524 LPMDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibhe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearseall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akazawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda I.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yanahara S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleishmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balasubramaniam S., Barth G.,
RA Blake J., Boffelli D., Bojunga N., Carinici P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereaux P.,
RA Nourjone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DB EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB33688.1; -.
DR EMBL; AK004883; BAB33641.1; -.
DR EMBL; AK004911; BAB33662.1; -.
DR MGD; MG1:95294; Egfr.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.8%; Score 1482.5; DB 11; Length 655.
Best Local Similarity 43.2%; Pred. No. 5.7e-104; Matches 276; Conservative 99; Mismatches 241; Indels 23; Gaps 7;

QY 11 LVALLPAGAA--STQVCTGDMKLRLPASPETHLDMLRHLVGGQVQGNLELYLPTN 68
DB 14 LITLCAAGALEEKVKQGTSTNRLTQIGTFEDHFLSLQRMVNNCEVLLGNLEITYQFN 73
QY 69 ASLSFLQDIQEVQYQIKANSKFIGITELQRLRIYRGTOLEFDNVALAVLNDGPNLNTTP 128
DB 74 YDLSPKTIQEVAGVLIATNTVERIFLENQITRGALVENTYALALSN----- 124
QY 129 VTGASPGGLRELOLSRLTEILKGVLIQRNPOLCYQDTILMKDI----FKKNQALATLI 184
DB 125 -YGNRTGRLPRLPMNLQELLIGAVRSSNNPILCNMTIQMRDIYQVAFNMSMDL--- 180
QY 185 DNNRSRAHPGSPNCKSGKCGESSEDDQSLRTVCGGCA-RKGLPLPDCCGCGCAAG 243
DB 181 -QSHPSGCPKDPSCPNCSWGGEENCQKLTIKLCAQCSHRGRGSPSCCNQCAAG 239
QY 244 CGPHSGSLACLHNHSGICELCPALVTYNTDTPFSMPPEGRYFGASCVACPYN 303
DB 240 CTGPRESDLVQKQKQDQATKQDCTCPPLMLYPTTYQMDVNPESKYFGATCYKCRNY 299
QY 304 LSTDVSGTLVCPPLHNGEVTADGTQRCCKSKPCARVCGLGNHLREYRAVTSANIQE 363
DB 300 VYTDHSGCVRAAGPDIYEV-EEDGIRCKCKCDGCRKCNIGIGEFKDTLSINATNIK 358
QY 364 FAGCKKIKGSLAFIPESPDGPASNTAPLOEOQVETLEITGYLYISAMPSPDLS 423
DB 359 FRYCTAIGDHLILVARKGDSFRTPLDRELEILKYVEITGIFLLIQWPNWTDLH 418
QY 424 VFQNLQVIRGRILHNGAYSLTLOGISLGLRLSRLRELSGLALIHNTLCPVHTVPM 483
DB 419 AFENLEIRGRTRKQGGQSLAVVGINTISGLRSLKEISDDDIVISGNRLCYANTINMK 478
QY 484 QLFPRPHQALLHTANRPDECEVGEGLACHQICAGHCKMGPRPTQVNCOSFLRGQEEVE 543
DB 479 KLFGRPNKTKIMNRAKDKAVNVNPNLCSSSEGGWGPBPCVSCQVNSRRECEVEK 528
QY 544 CAVLOGLEREYVNAHCLPCHPECPQNGSTYCGPADQCVACAHKDPFCVAPCPSG 603
DB 539 CNILGGEREFENSECQCHPECLPAMNITTCGRPDNDCIOCAHITDHPHCVKTCPPAG 598
QY 604 VKPDLSTYPMKFPDEBACOPCPINCTHSCVDLDDYGC 642
DB 599 IINGENNTL-VMKYADANNVCHLCANCTYGCAGPLOGC 636

```

Search completed: July 22, 2003, 09:00:06
 Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6847

Sequence: 1 MEALACRWGJLLALLPFGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 20604715 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6232	91.0	1259	6	018735 canis faml
2	3134	45.8	1209	1	09QX70 ratius norv
3	3106	45.4	1210	11	09EP98 mus musculu
4	2729.5	39.9	1137	13	09MEF6 gallus gall
5	2729	39.9	1165	13	09YH40 xiphophorus
6	2313	33.8	1328	5	09YH40 xiphophorus
7	2042.5	29.8	1433	3	09BIB9 fugu rubrip
8	1776.5	25.9	419	4	09UK79 mus musculu
9	1739	25.4	367	11	08R2X1 mus musculu
10	1720	25.1	729	15	08E712 avian rous-
11	1718	25.1	729	15	08E714 avian rous-
12	1697.5	24.8	412	4	08WV04 homo sapien
13	1637.5	24.1	962	15	08A895 avian eryth
14	1645	24.0	345	15	08A468 avian eryth
15	1497.5	21.9	655	11	09WVFS mus musculu
16	1481.5	21.6	643	11	09ERV6 mus musculu

17	1286	18.8	1193	5	09Y1X8 ephydaria f
18	1192.5	17.4	1368	5	023821 caenorhabd
19	1159	16.9	1717	5	026566 schistosoma
20	1125	16.4	527	13	090836 gallus gall
21	992.5	15.8	478	11	09ES50 ratius norv
22	941.5	14.8	599	13	09PSH2 gallus gall
23	906	13.2	165	4	014256 homo sapien
24	887	13.0	176	11	0923V5 ratius norv
25	806.5	11.8	346	13	09Y1X8 ephydaria f
26	778	11.4	435	5	08S2W1 ratius norv
27	754.5	11.0	311	13	09Y1X8 ephydaria f
28	746.5	10.9	1362	13	09PVZ4 xenopus lae
29	723	10.6	149	6	09B6G6 oryctolagus
30	721	10.5	331	4	09BUD7 homo sapien
31	716	10.5	1671	5	09N1V5 biomphalari
32	691.5	10.1	1368	13	08UW85 paraliichthy
33	687	10.0	1418	13	08UW86 paraliichthy
34	674.5	9.9	1369	13	08UW86 paraliichthy
35	667.5	9.7	1472	5	09U5A8 bombyx mori
36	661	9.7	1358	13	073798 xenopus lae
37	643	9.4	1412	13	08UW84 paraliichthy
38	633.5	9.3	1418	13	08UW83 paraliichthy
39	632	9.2	1245	13	09YGH8 scophthalmu
40	629	9.2	1371	11	09QVW4 ratius norv
41	614.5	9.0	2144	5	09VDM4 ratius norv
42	608	8.9	987	11	091YMO mus musculu
43	606	8.9	935	4	096135 mus musculu
44	605	8.8	987	11	09WVR2 mus musculu
45	587.5	8.6	1036	4	007912 homo sapien

ALIGNMENTS

RESULT 1	018735	PRELIMINARY:	PRN: 1259 AA.
ID	018735		
AC	018735		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Erdb-2.		
OS	Canis familiaris (dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yokota H.		
RT	"cDNA cloning of erdb-2 from canine mammary gland."		
RL	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB008451; BAA23127.1; -		
DR	HSSP; P1362; 1EGK.		
DR	InterPro; IPR002048; EF-hand.		
DR	InterPro; IPR000494; EGFR_L domain.		
DR	InterPro; IPR000719; Euk_Pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_Pkinase.		
DR	InterPro; IPR004019; YLP motif.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; Pkinase; 1.		
DR	Pfam; PF01030; Recep_L domain; 2.		
DR	Pfam; PF02757; YLP; 2.		
DR	ProDom; PD000001; Euk_Pkinase; 1.		
DR	SMART; SM00261; FU; 3.		
DR	SMART; SM00219; TYKc; 1.		
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
KW	ATP-binding; Transferase; Tyrosine-protein kinase.		
SQ	SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;		

Query Match	91.0%; Score 6232; DB 6; Length 1259;
Best Local Similarity	90.9%; Pred. No. 0;
Matches 1153; Conservative	40; Mismatches 59; Indels 16; Gaps 4;
QY	1 MELALCRWGLLLALLPGMASTQVCTGDMKRLPASPEHLMDLRLHLYGQCVVQGNL 60
DB	1 MELAAWCRWGLLLALLPGMASTQVCTGDMKRLPASPEHLMDLRLHLYGQCVVQGNL 60
QY	61 ELTYLPNTASLFLDIOEVGVYLIAHNOYRQVLOSLRVRTQOLFEDNYALAVLDNG 120
DB	61 ELTYLPANASLFLDIOEVGVYLIAHNOYRQVLOSLRVRTQOLFEDNYALAVLDNG 120
QY	121 DPLNNFNFTVSFMLRVKVSASHLE--QLRSLTEILKGVLIQRNPQLCYQDTILMKDI 178
DB	121 DPLEG-----GIPAPCAAGGRLRELQLRSLTEILKGVLIQRNPQLCHQDTILMKDV 172
QY	179 FHKNNQALTLIDITNRSPACHFCSPMCKSGSKGSESSDCSLRTVYAGGCAKCKPLP 238
DB	173 FHKNNQALTLIDITNRSPACHFCSPMCKSGSKGSESSDCSLRTVYAGGCAKCKPOP 232
QY	239 TDCHQEQACGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNPEGRYTFG 298
DB	233 TDCHQEQACGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNPEGRYTFG 292
QY	299 ASCVTACPFNYLSTDVSGCTLYVPLHNOEVTABDGTORCEKSPCARVCYGLAMEHLRE 358
DB	293 ASCVTSCPYNYLSTDVSGCTLYVPLHNOEVTABDGTORCEKSPCARVCYGLAMEHLRE 352
QY	359 VRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTPLPDELQVPELLEETGLYLT 418
DB	353 VRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTPLPDELQVPELLEETGLYLT 412
QY	419 SAMPSLPLSVFONLQVIRGRILHNGAYSLTLOGLSWLGRLSRLSGLALIHNT 478
DB	413 SAMPSLPLSVFONLQVIRGRILHNGAYSLTLOGLSWLGRLSRLSGLALIHNT 472
QY	479 HLCFVHTVPMDOFLFNPQOALLHTANRDEDCVEGGLAQHQLCARGHOMKPGPTQNCNS 538
DB	473 HLCFVHTVPMDOFLFNPQOALLHTANRDEDCVEGGLAQHQLCARGHOMKPGPTQNCNS 531
QY	539 QFLRQCEVEECRVLQGLPREVYNAHNLCPHPECOPONSVTCFGEADQCVACAHYKD 598
DB	532 QFLRQCEVEECRVLQGLPREVYNAHNLCPHPECOPONSVTCFGEADQCVACAHYKD 591
QY	599 PFCVACRCPGSKVPLSLMPIMKPFDEGACQCPINCHSCVDLDKCPAEOPLSPLT 658
DB	592 PFCVACRCPGSKVPLSLMPIMKPFDEGACQCPINCHSCVDLDKCPAEOPLSPLT 651
QY	659 SIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRRLLCETELVEPLTSGAMPNOAQM 718
DB	652 SIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRRLLCETELVEPLTSGAMPNOAQM 711
QY	719 RLKETEIRKVKVLSGAFGVYKICIMPDSENVKIPVAKIVLRENTSPYANKKILDEAV 778
DB	712 RLKETEIRKVKVLSGAFGVYKICIMPDSENVKIPVAKIVLRENTSPYANKKILDEAV 771
QY	779 VMAGVSPVSRLLGICLTSTVOLVTOAMPYGCCLLDVHRENRGLSGODLLNKMOMIANG 838
DB	772 VMAGVSPVSRLLGICLTSTVOLVTOAMPYGCCLLDVHRENRGLSGODLLNKMOMIANG 831
QY	839 MSYLEDLVLRDLAARVNLVKSNNHVKITDFGLARLLDIDETEHADGGKVIDIKMALE 898
DB	832 MSYLEDLVLRDLAARVNLVKSNNHVKITDFGLARLLDIDETEHADGGKVIDIKMALE 891
QY	899 SILRRFTTHOSDWSYGTVMELTFGAKPPDGLPAEIRIDLEKGRLLPOPICTIDVY 958
DB	892 SILRRFTTHOSDWSYGTVMELTFGAKPPDGLPAEIRIDLEKGRLLPOPICTIDVY 951
QY	959 MIMYKCMWIDSECPREFELVSESRMAPDPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1018
DB	952 MIMYKCMWIDSECPREFELVSESRMAPDPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1011
QY	1019 DMGDLVDAEYLVPQGGFFCPDPAFGAGMWHHNRSSSTRSGGDLTLGLFSESEAR 1078

DB	1012 DMGDLVDAEYLVPQGGFFCPDPAFGAGMWHHNRSSSTRSGGDLTLGLFSESEAR 1078
QY	1079 SPLASEGAGSDVPFDGDLGMAAGKGLSLPHDSPLQRYSEDPVLPSETDGVAPLT 117
DB	1072 SPLASEGAGSDVPFDGDLGMAAGKGLSLPHDSPLQRYSEDPVLPSETDGVAPLT 117
QY	1139 CSPQPEYVNOQDVNRQPPSPREGEPLPAARPAATLER-----ATTLSBCKNGVVKVDFAF 117
DB	1132 CSPQPEYVNOQDVNRQPPSPREGEPLPAARPAATLER-----ATTLSBCKNGVVKVDFAF 117
QY	1194 GGAIVENPEYLPQGGAAQPPHPPAFSPAFNTLYYMODPDERGAPSTFGPTAENPE 127
DB	1192 GGAIVENPEYLPQGGAAQPPHPPAFSPAFNTLYYMODPDERGAPSTFGPTAENPE 127
QY	1254 YLGLDVPV 1261
DB	1252 YLGLDVPV 1259
RESULT 2	
ID	090X70 PRELIMINARY; PRT: 1209 AA.
AC	090X70;
DT	01-MAY-2000 (Tremblrel. 13, Created)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Epidermal growth factor receptor.
GN	EGFR.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_Taxid=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RX	MEDLINE=9025888; PubMed=2342466;
RA	Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
RA	Earp H.S.;
RT	"A truncated, secreted form of the epidermal growth factor receptor is encoded by an alternatively spliced transcript in normal rat tissue."
RM	Mol. Cell. Biol. 10:12973-2982(1990).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RA	Petch L.A.;
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER; TISSUE=LIVER;
RA	Guttridge K.; Dawson T.L.; Earp H.S.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; M37394; ARI1408.1; -
DR	HSSP; P11362; IREG.
DR	InterPro; IPR000494; EGFR_L domain.
DR	InterPro; IPR000719; Euk_Pkinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_Pkinase.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF00069; Pkinase; 1.
DR	Pfam; PF01030; Recep_L domain; 2.
DR	PRINTS; PR00109; TYRKINASE.
DR	Prodom; PDC00001; Euk_Pkinase; 1.
DR	SMART; SM00261; FU; 3.
DR	SMART; SM00219; TYRK; 1.
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR	PROSITE; PS00109; PROTEIN KINASE DOW; 1.
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW	ATP-binding; Receptor; Transferrase; Tyrosine-protein kinase.
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEERF6C1B7773 CRC64;

Query Match 45.8%; Score 3134; DB 11; Length 1209;
 Best Local Similarity 49.7%; Pred. No. 7; 6e-229;

Matches 638; Conservative 166; Mismatches 359; Indels 120; Gaps 27;

```

QY 3 LAALCRWGLLALPPGA-ASTOYCTGDMKLRLPAPBETHLMDLRLYGGCQVQNGLE 61
Db 15 LAALCAAG-----GALBEKKVCOGTSNRLTOLGTFEDHFLSLQMFNNCVAVLGNLE 66
QY 62 LTYLPTNASLSFLDIOEVQGYVLIHNOVROVELQRLRTVGTOLPEDNYALAVLDNG 121
Db 67 ITTVQRNYDLSFLKIQEVAGYVLIANTVERIPLENLQIIRGNALYENTYALAVLSN-- 124
QY 122 PLNNFNNTVSFWLRVPKVSASHLEQLRLSLELLKGVLIQRNPOLCYQDTIIMKQIFHK 181
Db 125 -----VGTNKTGLREL-PRNNLOELIIGVRFSSNPNILCNMTIQRDIIV-Q 169
QY 182 NNOALALTIDITNRS-RACHPCSPMCKSGRCWGESSEDCQSLTFTVCAGCA-RCKEPLFT 239
Db 170 DVFLSNMSMDYQRLHITGCPKCDPCPNMGSCGRREBNOCQTLTKICACQCSRRCRGRS8 229
QY 240 DCHEQCAAGCTGPKHSDCLALHPNHSIGICEHPALVTYNTDTFESMNPGRYTFGA 299
Db 230 DCHNQCAAGCTGPRSDCLVCHRFDEATCKDCPPLMYPPTYQMDVNPBGKYSFGA 289
QY 300 SCVTACPNYVLTVDGSCITVCPJLHNOEVTAEADGTORCEKSCPRCAVCYGLMEHLREY 359
Db 230 TCVKCKPRNYVTVHSGCYRACGPDYEV-EDGVSKCKKCDGCRKVCNGIGIGEFKDT 348
QY 360 RAVTSANIQEFAGCKKIFGSLAFIPESPDGDPASNTAPLOEOLQVFEETLEITGYLIS 419
Db 349 LSINATNIKHKFKYCTAISGDLHIIPVAFKGDSPFRTPPLPRELEILIKTKETIGLLIQ 408
QY 420 AMPDSLPLSVFQNLQVGRILHNGAYSLTLQGLGSMGLSLELSGLALIHNNH 479
Db 409 AMPEWTDLHAFENUEIIRGTRKHGQPSLAVGLNITSLGLSLKEISDGVITISGNH 468
QY 480 LCFVHTVPMQDLFRNPQALHTANREDEDECVGGLACHOLCARHCWGPPTQVCNCSQ 539
Db 469 LCVANTIMWKKLFTGPNQKTKIMNRAEKOKATNNHCNPLCSSEGWGPEDVCSCQN 528
QY 540 FLRGOCEYECRVLCQPREVYANRHLCPHPECQOPNGSVTTFGEADOCVACAHYKP 599
Db 529 VSRGECVCKNILEGEPREVENSECIOCHPELPTQNMNITGTGGRPNCKICAHYVVG 588
QY 600 PFCVACPSGVKPLSLYMPIKFPDEEGACOPCINTHSCVDLDRKGPABGRASFP-LT 658
Db 589 PHCVKTCPSGIMGNNLT-VWKFADANNVCHLANCTYCAAPGLKC--QOPSPKLP 645
QY 659 SIVSAVVGILLIVLVGVVFGI-LIKRQOKIRRYTMRLIQETELVEPLTPSGAMENQAQ 717
Db 646 SIATGIVGGLFIVV-VALGIGLFMRRLQVRRKTLRLLQERELVEPLTPSGEARNQA 704
QY 718 MRILKETELRKVXLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTPKANKETLDEA 777
Db 705 LRIIKETEFKKIKVLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTPKANKETLDEA 764
QY 778 YVMAGVSPYVSRLLGLCTSTVQLVTQMLPYGCLLDHVENRGRGLSDLLNMCWQIAK 837
Db 765 YVMASVNPVHCRLGLCTSTVQLVTQMLPYGCLLDYVREHKDNGISQYLLNMCWQIAK 824
QY 838 GMSYLEDVRLVHRDLAARNLVKSPNNVKTDTSGIARLLDIDTEYHAGSKVPIKMAAL 897
Db 825 GSNYLEDERRLVHRDLAARNLVKIPQHVKTIDTGLKLLGAEEKEHYHAGSKVPIKMAAL 884
QY 898 ESILRRRFTHQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGEELPPPICTIDV 957
Db 885 ESILHRIYTHQSDVSYGVTVWELMTFGSKPYDGIIPASEISILKGEELPPPICTIDV 944
QY 958 YMIWVKCMMIDSECRPRFELVSEFSMADPRRFVYIQ-NEULGASLSDSTFRSLIE 1016
Db 945 YMIWVKCMMIDADRPRFRELILFSSMADPRRYLVLIQDDEWMLPSPJTDNFPYALME 1004
QY 1017 DDDWGLDVDAEEXYLPVQGFCCPDPAFGAGMWHHRSSSTRSGGDLTLGLESEBEA 1076
Db 1005 EEDWEDVVDADDEVILPQGGFF-----NPSST----- 1030

```

```

QY 1077 PRSPLAPEGAGVDFDGLDGMGAKLQSLPTHDPSPILQRYSEDPVPLPSET--DGYV 113
Db 1011 SRTPLSLSLANSN-----SSTVACINNGSCRYKEDFLQRYSSDPVSLTEONIDTFL 106
QY 1135 APLTCSPOPEYVNPDPVPPQPPSPREGFLPAAPAGATLERRATLSFGKNVAKVDFAFG 115
Db 1087 -----PVPEYINQ-SVPKRPAGSVQNPVYHNGPLNH-----APGRRLHQN--PHS 112
QY 1195 GAVENPEYL-TPQGAAPQPHPPAPSPARDNLYWMDQ-----DP-----PERG 123
Db 1130 NAVSNPEYLNTAQ-----PTCLSSGPDSSALWIQKSHQMSLDNPDYQODFFPKXA 118
QY 1238 APPSTFKGTPAENPEYLGIDVP 1260
Db 1181 KPNGIFKG-PIAENAEYLRVAPP 1202

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVETAC;
RA Retter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Retter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; IFGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; Egrf.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_Chr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PFO1030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SMO0261; FU_5.
DR SMART; SMO0220; S_TKc; 1.
DR SMART; SMO0219; TYKc; 1.

```


Query Match 39.9%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity 47.3%; Pred. No. 3,66-198;
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

167 LCYQDTILMKDIFHKNNQALALIDTNSRACHCSPKCKSRWGSSSDQSLRTTVC 226
3 LCFADTILHMODIVRNPMASNFTLVFTNSSSGCRCHSCGTG-RWGMPTENHCOTLTCTVC 61

227 AGGC-ARCKGRLPTDCHCEQCAAGCTGPKHSDCLACHFNHSGICELCPALVTYNTDTF 285
62 AQCQGRGRCYGPVSDCCRECAAGCGSGKXDTDCACNPNFNSGACVTCQQTPTVYNTTF 121

286 ESMNPBGRITFGASCVTACPNYITSDVSGCTLVCPHNOEVTAEQTPRCCKSPCA 345
122 QLEHNHNAKYTYGACVCKCPHNFTV-VDSSCVACPCSSKMEV-EENGIMCKRCPCTDICP 179

346 RVCYGLGMEHLREVAVTASANOEPAGCKIFGSLAFIPESPDPPASNTAPLOPELOV 405
180 KACDQIGTGLSVAQTVDSNIDKFINCTKINGNLILVTGIGHDPYHTTAATNPERKNT 239

406 FETLEITGYLYISAMPDLSPLDSVPONLQYIRGRILHNGAVSLTGLGISMVGLRSIR 465
240 FQTVREITGYLNTIOSWPEMNTDFRVFSMLVTIGRALYSGSLILLIKQCGITSLQPSLK 239

466 ELGSLALIHNTHLCPVHTVPMQDLFRNPHQALHTANRPDEQCEGLACHQLCARGH 525
300 QISAGNIYITNSNLCYHIVNMTLSLSTPSQKTVHNNKAENCTADGACNELSSDG 359

526 CWGPGPTQCVNCSQFLRQCEVCEECRYLOGL-FREYVNAHRLCPHPECPQ-ONGSVTCFQ 584
360 CMGPGPDQCLCKKPIRGRTCTESCNLYDGEFRFANGSVCMEDPCCEKEDNMTICG 419

585 PEADQVCAHYKXPPFCVACBSGVKPDLSYMIKFPDEBGAQCPQINCTHSCVDLD 644
420 PGPHTCKCFHFKQGPCVCEKCPDGLQANSF--IFKVADEDECHCHENCTQCGRPA 477

645 DKGC-----PABORASPLTSIVSAV--GILVVLVGVVFGILIRROQKIRKXT 692
478 SHDCIYVPMTRQSTLPHAR-TLP--IAAGVIGLFTIVMGLTFVAYVRKSIK-KKA 533

693 MRLLQETELVEPLTPSGAMPNOQMRLKETELRYKVYLGSAFGTYVKGWIPDGENV 752
534 LRRL-ETELVEPLTPSGAMPNOQMRLKETELRYKVYLGSAFGTYVKGWIPDEGTV 592

753 KIPVAIKVIRENTSPXANKELDAAVYAGVSGVYSRLGILCTVQLVQLMRYGL 812
593 KIPVAIKILNETTPKXANVEFMDALIMASMDHPLVRLGVLCSPTQLVQLMRYGL 652

813 LDHVENRGLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSNNHKTIDFGL 872
653 LDYVHEKDNIGSOLLNMCVQIAKGMVLEERRLVHRDLAARNVLYKSPNHKIIDFGL 712

873 ARLLDIETEHAGGVPIKMALESILRRRTQSGDWSYVYVWELMTFGAKPYDGI 932
713 ARLLGDEKEYNAGGMPIKMALECIHYRKTQSDVWSYVYVWELMTFGAKPYDGI 772

933 PAREIPDLEKGERLPOPICTIDVYVIMVKCMWIOSECPREIREIVSEFSRARDPORF 992
773 PTREIPDLEKGERLPOPICTIDVYVIMVKCMWIDADSPKKELAASFMAADPORF 832

993 VVIONED-LGPASLSDTFYRSLLEDDMDGLVDAEYVLPQGFPCPDPAAGAGVNH 1051
833 LVIOGDDRMKLPSPNSKPFQNLDEDEDLDMDAAEYVLP-QAFNIPPIYTSRTIDS 891

1052 RHSSSTRSGGDLTLGLEPSEEEAPRS--PLAP-SEGSGSDVFDODLGMGAKLQSLP 1108
892 NRNGFYVRDGYAAEGCV-PMYRABGCIPEAPVAGATAELEFEDTCNGTLRKQVATL 950

1109 THDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVNPDPVRPQPSREG 1161
951 AKEDSSTQRYSAPTVFIPIERVIRGELDEDEGYTMPRDXPKTDYANPVENFVSRKNG 1010

1162 PLPAA-RPAGATLERAKTISFGKGVVQVVF-----AFGGAIVENPYLTPQGAAPQ 1212

DB 1011 DLQADNDEYHN-----ANQCPKXADEYVNPPLYNFPAITLENAEYL-----K 105

QY 1213 PAPPAPSPADNLYYMDQDFPERGA--PSTFKGPT-----AENPEYL 1235

DB 1056 NNLPEKAKKARDNDPDYWNHSLPPRSTLQHPDYQEVSTKYFKONGRIRPIVAENPEYL 1114

RESULT 5

QYH40 PRELIMINARY; PRT; 1165 AA.

AC QYH40; 01-MAY-1999 (TREMblrel. 10, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Receptor tyrosine kinase proto-oncogene.

GN XMRK.

OS Xiphophorus xiphidium.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorphia; Acanthopterygii; Percormorpha; Atherinomorphia;

OC Cyprinodontiformes; Poeciliidae; Xiphophorus.

OX NCBI_Taxid=8086;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RIO PURIFICATION;

RX MEDLINE=98241172; PubMed=9582016;

RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,

RT Altschield J., Schartl M.

RT "Activation of the Xmrk proto-oncogene of Xiphophorus by overexpression and mutational alterations."

RL Oncogene 16:1681-1690(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=RIO PURIFICATION;

RA Schartl M.

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; U53471; AAd10500.2; -.

DR HSSP; P11362; IFGK.

DR InterPro; IPR000345; CytC_heme_bind.

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR000719; Euk_Dkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001899; Gram_pos_anchor.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Kinase; 1.

DR Pfam; PF01030; Recept_L_domain; 2.

DR PRINTS; PRC0109; TYRKINASE.

DR PRODOM; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TYRK; 1.

DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 2.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR KAT-binding; Kinase; Tyrosine-protein kinase.

SEQUENCE 1165 AA; 129614 MW; 7E7EB38D8771A7AE CRC64;

Query Match 39.9%; Score 2729; DB 13; Length 1165;
Best Local Similarity 45.5%; Pred. No. 4,1e-198;
Matches 583; Conservative 163; Mismatches 382; Indels 154; Gaps 30

1 MELALCRMGILLALLPPG-AAST-----QVCTGTDMKRLPASPETHLDMLRHLYQGCY 55
4 LELLE-----LILLILISIGRCSTIDBRKVCQGTNSQMNM---LDNHLYKMKMYSGCNV 56

QY 56 VQGNLELTYPPTASISFIOODQEVGYLYLIANQVQVPLQRLAVGTQLFEDNYALA 115

DB 57 VLENLETTQEQDLSFQSIQIEVGGVYLLIMNEVSTIPLVNRLIRGONLIEGNFTLL 116

QY 116 VLDN--GDPLNNFNFTVSWFLRVKVASHLEQRR--SLTEILKGVLIQNPQLCYQD 171

```

Db      117 WMSVQKRPSS-----PDYVGVGLKQQLSNLLEILSGVKYSHNPLCNVE 163
QY      172 TLKMDIFKXNQLALLIDINRSRACHPCSPMCKGRCSWSESSEDCSLTRTVACAGC- 230
Db      164 TINMWDIVDKTNSPTMNLIPHAFERQCKCDPGCVNCSMAFGPGHCQKFKTLCAECN 223
QY      231 ARCKGFLPTDCCHCCAGCTGPKKSDCLACLFHNSGICLHRLPALVTYNTDTFESMPN 290
Db      224 RRORGPPIIDCCNEHAGGCTGPRATDCLACDFDDDTCKDTPPKXIDIVSHQVVDN 283
QY      291 PEGRYTFGASCVTACPYNYLSTDVGSCTLVCPHNOETAEDGTQRECKSKPCARCYG 350
Db      284 PNKXTFGAACVKEGCSNVVTE-GACVRSAGALEVD-ENGKRSCKPCDGVCPKXCDG 341
QY      351 LGMEHREVRATSNICQFAGCKKIFSSLAFLPESFGPDASTAPLOPEQLOVFETLE 410
Db      342 IGISLSTIANVSNIGFSNCTINDIILNRSFEGDHYKIGPMDEHMLNLTIVK 401
QY      411 EITGYLISAMPDLPDLVFNOLQVIRGRILHNGAYS-LTLOGLSIWLGRSLRELGS 469
Db      402 EITGYLIVIMWPEMNTSLSFONLEIIGRTTFSGSFVYVQVSHLQWGLSLKEVSA 461
QY      470 GLALIHNTLHCFHTVPMDOLENNPHOALHTANRPEDECVEGELACHOLCARGHMG 529
Db      462 GNVILKNTPOLRYASTIMWRRLFSEDSQIEYDART-----ENQTCNNECSHDGCMGP 514
QY      530 GPTGCVNCGQFLRGEQCEVECEVLOGLEPREVYNAHCLPCHPEQOPQNGSVTCGPEADQ 589
Db      515 GPTMCVSLHVDGRGCVASCNLLOGEREQVDRGCQCHQBELVYDTSITCGPEPAN 574
QY      590 CVACAHKDPPECVARCPGSKVPDLSTYPIWKFPEDEGACPCPINCTHSCVDLDDKCP 649
Db      575 CSKCAHFDGPOCIIRCPHGMIGDGL-IKXADKMKQCCPCHQNTQCSGSGSGCR 633
QY      650 AGRASPULSTVSAYVGLLVVGVGGLIKRQCKIKRYTKRRLLOETELVEPTLPS 709
Db      634 GD-IVSHSLAVGLVSGLLITVYALVLVLRRLRIK-RRKTIIRLLQKEKEVEPTLPS 691
QY      710 GAMPNQAQMRILKETELRKKVYLSGAGFYKGIWIPDGENYKIPAIKIVLRENTSPKA 769
Db      692 GQAPNQAFLRIKKEFEKKDVLGSGAGTYKKLMPDGENIRIPAIKIVLEAISPXY 751
QY      770 NKEILDEAYVMAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYENRKGSLGSDLL 829
Db      752 NOEVLDEAYVMAVDHPHYCRLLGICLTSAVOLVTOMLPYGCLLDHYRQHGERICQWLL 811
QY      830 NMCQOIAKGMSTLEDRVLVHRDLAARVLYKSPHVKITPGLARLLDIDETRYHADGK 889
Db      812 NMCVOIAKGMNLEBRHLVHRDLAARVLYKSPHVKITPGLARLLDIDETRYHADGK 871
QY      890 VPIKMALESILRRRFTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLQ 949
Db      872 VPIKMALESILQWYTHQSDVMSYGVTWELMTFGSKPYDGIIPAKETIAVLNGERLQ 931
QY      950 PICTIDVYMIWVCOMMIDSECPREFELVSEBRMADQRFVYVIONEDLGASPLDST 1009
Db      932 PICTIEVYMIWVCOMMIDSPSRPREFLVSEFQMARDSRYLVIG--NLPSPSDR 988
QY      1010 FYRSLLEDDMDGLVDAEYLVPQCGFCPPDPAAGAGMYHHRSSSTRSGGDDLTLGL 1069
Db      989 LFSRLSSDD--DIVDADEYLL-----RYKRIN-RQGS----- 1018
QY      1070 EPSEEARPSPLASSEGAGSDVFDGDLGMKAAGLQSLPHDPSLQRYSEDPTV-PLPS 1128
Db      1019 -----EPCTIPNGH-----PVRENSIALRYISDPTQVALLER 1049
QY      1129 ETDGVAPLTCSPQPEYVNDPVRPPQ-----PSPRE-----GPPR-AARAGATLER 1175
Db      1050 DLDDG-----EYNQPGSETSSRLSDIYNPNVEDLTDGKGVYSSQCAETNFSR 1099
QY      1176 AKTISPGKGVVAVFAFGAVENPEVLTTPGGGAAPRPPAPFAFDLVYVWDDPPE 1235

```

```

Db      1100 PEYLTNQNLSL----PLVSSSGMDPDY----QAG-----YQAAF-----LPQ 117
QY      1236 RGAPPSTFKGTPTAENPEYIGL 1257
Db      1135 TGAITGNMGFLPAENLEYLGL 1156

RESULT 6
P79754
ID P79754 PRELIMINARY; PRT: 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMBlrel. 03. Created)
DT 01-MAY-1997 (TREMBlrel. 03. Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21. Last annotation update)
DE Etd33.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxId=31033;
RN (1)
RP SEQUENCE FROM N. A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelliner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wrt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P1362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_Dkinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

Query Match 33.8%; Score 2313; DB 13; Length 1328;
Best Local Similarity 40.4%; Pred. No. 1,9e-166;
Matches 523; Conservative 154; Mismatches 417; Indels 200; Gaps 31

9 WGLLALILPP--GAASQ-----VCTGDMKLRLPASPEHLDMLRHLYQCCQVVGNTLEI 62
Db 4 WRLILMCVASSLRASSQTOEAVCPGTQNGLSSTGSENGYNLMKRYKCEIIMNLEI 63
QY 63 TYLPTNASLSLQDIOGVQGVLIANQVRQVPLQRLIRYRGQLFEDNYALAVLDNGDP 122
Db 64 TQIESNMDPSFLKTIREVTVYLIAMNHQPEIPLGQLRVIRGNSLYERFALSV----- 117
QY 123 LNNFNNTFVSWLKVPRVYASHLQ--LRSLTEILKGVYIQGNPQLQVQDTILMKDIFH 180
Db 118 -----FLNYPDGPSSGLNQGLAMNLTILDGQVQIINNKYLRIGPVYVMDIT- 165
QY 181 KNNQLATLIDITNRSRACHPCSPMCKGRCSWSESSEDCSLTRTVACAGC-ARCKGFLPT 239
Db 166 RNNDAPIEQNGRGVCH--KSC-GNYCWPBGKQCCQLLTITVCAPQCNDRCFGTSR 221
QY 240 DCCHEGCAAGCTGPKKSDCLACLFHNSGICELHRCALVTYNTDTFESPNPEGRTFCA 299
Db 222 DCHTECAAGCKGLDTPCACLFLNDSGACVAPQCPQTLIYNKQTFQMETNPAKYQYS 281
QY 300 SCVTACPYNYLSTDVGSCTLVCPHNOETAEDGTQRECKSKPCARCYGIGMEHRE 358
Db 282 ICVQGCCTHVV-VDGSCVSVCPRPDKEV--ERSGRQCELSGGLCPKXCEBTGAE---Q 335

```

```

QY 359 VRAVTSANIQEFAGCKKIFGSLAPFESFDGDPASNTAPLOPOLQVETLEITGYLYI 418
DB 336 ROTVSSNIDSFINTKIQGSLHFLVTLGDDPKVNPIDAKLEVFRTVREITDILNI 395
QY 419 SAMPSLPLDLSVFOVLQVIRGRILHNGVSLTLOGLSIMGLRSLFELSGSLALIHANT 478
DB 396 QSWPELNDLSVFSLLTIGRSLFKRSLVMWRIPLTSLGRLSLREISIDGVSYSQNA 455
QY 479 HLCFVHTVPMDOLFENPH-QALLHTANPEDECEGEGELACHQLCARHCGPPTQCVC 537
DB 456 HUCYHTVNMVTLQFGRSVRANSLNSNRPMAECVADRVDPLCSGSCGPGPDQLSC 515
QY 538 SCLFAGQCEVECKVLOGLPREYVNAH-CLPCEPQONGSVTCGFPADQCVACAHY 596
DB 516 RNYSHGTGVAGCHENSIPREFAGLNGVCVAHBECKPQGXKASCTGFPADCEMACTKF 575
QY 597 KDPFCVACPSGVPRDLSYMPIMKFPDEGACQPCDINCHSVDJDDGCAEGRASP 656
DB 576 RDGPYCMSSCPAGVN-DEKGLIFKFRBEHCPCQNCCTQCGSGGLNDC--LEAAH 631
QY 657 LTVSVAVGLLVVLGVF-----GILIKRQCKIRKYTWRLLOTELEVEPLTS 709
DB 632 LTSSGQITGIALGPAGLIFCLVLFGLMLYHGLAIRRRAMRRLYEGESFEPLGP- 690
QY 710 GAMPQACMRILKETEELKVKVULGSGAFGYKGIWIPDENKIPALIVYLENTSPKA 769
DB 691 GEKGTAKARILKSDLKIKPLSGVGVISKGFMIPEBETVXIPALITDSSGRQT 750
QY 770 NKEILDEAVVAGVSPVSRLLGICLTSTVQTLQMPYGLDLHRENRGLSGQDL 829
DB 751 FTEITDHLMSGLDHPYIVRLGICPOTGLVTLQSHSGSLHEHROKHTLSDQRL 810
QY 830 NMCQIAGKMSYLEDVRLVHDLAAVVLKSPHVKITPGLARLLIDETVHADGK 889
DB 811 NMCQIAGKMSYLEDVRLVHDLAAVVLKSPHVKITPGLARLLIDETVHADGK 889
QY 890 VPIKMALESILRRFTHQSDVMSYGVTWELMTFPAKPDGIPAREIPPLEKGERLPQ 949
DB 871 TPIMMALESILRRFTHQSDVMSYGVTWELMTFPAKPDGIPAREIPPLEKGERLPQ 930
QY 950 PPICTIDVYMTVMVKCMWIDSECRPFRELVSFGRMARDPQRFVIONEDLGPASPLDST 1009
DB 931 PAICTIDVYMTVMVKCMWIDSECRPFRELVSFGRMARDPQRFVIONEDLGPASPLDST 980
QY 1010 FYRSLDEDDMDGLVDABEYLVPOQGFCCDPAPAGACGMWHHRSSSTRSGGDLTGL 1069
DB 981 -----EDSGMGEFL-----RGSER--GLLEADL 1002
QY 1070 EPSEEAAPRSPPLABSEGAGSDVPDGLGMC--AAKGLQSLPTHDPSPLO----- 1116
DB 1003 EEDDEE-----GLDRAFATPBLQPSPEWSTSPQINSYMMVMTQL 1041
QY 1117 RYSEDPYVPLSETDGVAVPLTCSPO-ETVNO-----PDVROPSPSPREG 1161
DB 1042 RYD-----FAVSGGHHGYLPMSPSPVDITRQLWYQSRSLSVRTLPPRSARRSSRA 1095
QY 1162 PL--PAARPAATLEPAKTLSPGKNGVAVKDFAFGCAVENPEVLTPOGGAAPQHPHAF 1219
DB 1096 ELCEGAGCAGIFVVR-----FGSRGN-----FQGG----- 1122
QY 1220 SPAFDNLVYMDQDPPEKAPPSFTFKGTPTAENPE 1253
DB 1123 -----QQRKLSTASSPSSFKTWADEDE 1146

```

```

GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxId=7165;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN:SU4.
RA Lycet G.U.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RT Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ301655; Genbank: CAC35008.1;
DR HSB: P11362; IFCG.
DR InterPro: IPR000345; CyfC heme bind.
DR InterPro: IPR000494; EGFRL domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_Thr_Pkinase.
DR InterPro: IPR01245; Tyr_Pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SMO0261; FU; 7.
DR SMART: SMO0220; S_TKC; 1.
DR SMART: SMO0219; TYR_KC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR KX Receptor.
FT NON TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D8896724F07 CRC64;

Query Match 29.8%; Score 2042.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 7.6e-146;
Matches 474; Conservative 197; Mismatches 394; Indels 385; Gaps 37

QY 26 CTGTDMKLRLPASETHLDMRLHLYGQCCVYVQCNLELTYPTNASTSFIQDIOEYGVYL 85
DB 1 CTGNGKMSVPANNEYHYKNLRRRYTNCYVDNLELITWQNTITDNLFIQHIREVYGYVL 60
QY 86 IAHQVQVPLQRLRIYRGTLQFEDNYALAVLDNGDPLNNFNFTVSFMLRVKVSASHL 145
DB 61 ISLYDLPQVILPRQIRGRTRK-----LKMGEAYGLF-----VSFSHM 101
QY 146 E--QLRSLTEILKGVULIQNPQLCYQDITLAKDI-FHKNNQALTLIDTNRSPACHPS 202
DB 102 NTLLEPALRDILGSGVGFNNVNLCHMKSLNMBEILLAPQTSQVTFNFSSEPRVCPCH 161
QY 203 PMCKGSRGMESSBCCSLTRTYCAGCA--RCKGFLPTDCCHEGCACTGPKSHSDLA 260
DB 162 PSCGCVG-CWEGEAGNCRKSLKNCSPQCSGGRFGFKRPECCLFLGAGCTGPTQSDCLA 220
QY 261 CLAHNSGICELHCPALVYNTDIFESMPNPEGRYTFGASCYTAECYNYLSTDVSGCTLV 320
DB 221 CKNFYDDGVCKQCCPQIYNFNFMPEPDPKAYAGATCYVRKP-ETLLDNGACVYRK 279
QY 321 CPLHNGEVLTEDEGTQCEKSKRCACAVCYGLGMEHLREVRATVTSANIQEFAGCKKIFGL 380
DB 280 CPKGRKPFQNSE-----CVPKGVCPKTCGEGIVH-----SDNIGYKQCTTIEGSL 326
QY 381 AFLPESEFDGPASNT-----APLQPELOVFTLEETLEITGYLYISAMPDLSPLSVFO 432
DB 327 ELIDQSFQGVQYVTFNSFGPRYIKIDPRLLEFVSIVKXITIGFINIQAHHRFTLLNFR 386
QY 433 NLQVIRGRILHNGAY-SLTLOGSIGIWLGLRSLRELSGSLALIHANTLCFVHTVPMDQL 491
DB 387 NLEVYGGRLKENLFAVYIVKTSLSKSLKIKRVNSGSIVILENSDLCFVEDIDWSEI 446

```

RESULT 7
 Q9BIH9 PRELIMINARY; PRT; 1433 AA.
 AC Q9BIH9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative epidermal growth factor receptor (Fragment).

```

QY 492 FRPHQALHTANREDECEVGEGLACHQLCARHCQWCPPTQVNCQPLRGQCEVECR 551
DB 447 KSSDHEWVWQKRNATCHEGSECESSQCSKAGCWGPQCECKXVKKXGKCLDSCK 506
QY 552 VLOGLPREV-VNARHCLPCHPEQCFQNGSVTCFGEADQCAAHYKDPFCVAPCP--- 607
DB 507 ---SLPRLSYVDSKTCGDOHCECKD---FCYGNEDNCSCNNVADGRFCVACEPTTK 558
QY 608 -----SGYKPLSLNMPMKFPD----- 624
DB 558 HAMNGTCINCHKTCVCGCRGPRDTIAPDGCISCDAIIGSDAKIERCLMKDESCPDGYSD 618
QY 625 ---EEG----- 627
DB 619 VYLOEEGFLKOLSGKAVGRKCHPRCKCTGYGPHQFCQECTGYKKGEGCEDECPDQFYA 678
QY 628 ---ACQCPINCT-----HSCVDL-----D-----KCGPAEO----- 652
DB 679 NEETRICLPCHOEGCGCHGLDHDHECNLKLFGGDPYDNTTFTCVSNCPASHPYKRP 738
QY 653 -----RASPLTSYSAVAVGILLVVLGVVFGI---LTKRQOKIRK 690
DB 739 QEAGKIGPYCASDWSQSGRIEPCQVXIVWGSVMALLLCVVFGINFVLFSRHAKKDA 798
QY 691 YTMRLLOETELVEPLTPSGAMPNQAQMRILKETELRKXVVLGSGAFGYKGIWPDGE 750
DB 799 VKMTALAGCEDESEPLRPSNVGPNLTKLRIKEAIRGVLGAGARGVFKGWMEGE 858
QY 751 NVKIPVAIKVLENTSPRANKEILDEAVVAGVSPVYSRLGLCLSTQVLTQOLMPYG 810
DB 859 SVKIPVAIKVLENTSPRANKEILDEAVVAGVSPVYSRLGLCLSTQVLTQOLMPYG 918
QY 811 CLLDHVRENGRLSGQDILLNMCQIAKMSYLEDEVRLVHRDLAARVLYKSPNHWKITDF 870
DB 919 CLLDVYRNKKKIGSKALLNWSQIARGMAYLEBRRLHRLDLAARVLYKSPNHWKITDF 978
QY 871 GLARLLDIDETEYHADGKVPILKMMALLESILRRFTQSDVWSGYVWMLMFEKAPYD 930
DB 979 GLAKLLDSDSEYRAAGKMPILKMMALLESILRRFTQSDVWSGYVWMLMFEKAPYD 1038
QY 931 GIPAREIPDLLEKGERLPQPICTIDVYMIWVKWMTIDSECRPFRELVSEFSMARDPQ 990
DB 1039 NVPAKDVPELIEIGHKLPPQDICSIDVYCIILSCWLDADARFTFKQIATFAEKARDPQ 1098
QY 991 RRVVIONEDLGASLDSTFYRSLLEDODGDLV----- 1024
DB 1099 RYLMV-----PGDKFMRKLPSTYNODEKDLIRTLAPVAAAAAAGASNVDPVS 1149
QY 1025 ---DAEYLVLPQOGFFCPDPAFGAGWVHRHRSSTRSGGDLTLGLEPSEEEAPRS-- 1079
DB 1150 TIAETDEYLOPKTRSIIMLPQPSA-----VEBS-DEMKSLR 1185
QY 1080 ---PLAP---SEGAGSVYFGDGLCMGAAGLQSLPTHDSPLQRYSEDPVLPSEETDG 1132
DB 1186 YCKDPLKPDDETDGKXEV-----GVGGIR-----LNLPLDEDD 1219
QY 1133 YVAPLTCSPOPEYVNPQVPRQPPSPREGPLPAARPAATLERAKTLPCKGNVAVDVA 1192
DB 1220 YLMP-TCOSQ---NQS-----TPG---YNDLIG 1240
QY 1193 FGGAVENPEYL-----TPGGAAPQPPPPAFSPAFDNLYWDDQPERGAPSTF 1243
DB 1241 VPAVDNPEYLMGSTOALAGLAQSGMG--PHTP-----PPNTP 1277
QY 1244 KGTPTAENPE 1253
DB 1278 NGMPTHQHSQ 1267

```

```

DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.",
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN (2)
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SMO0261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEGCBE347E2D030C CRC64;

Query Match      25.9%; Score 1776.5; DB 4; Length 419;
Best Local Similarity 84.4%; Pred. No. 1,9e-126;
Matches 341; Conservative 5; Mismatches 37; Indels 21; Gaps 4

QY 1 MELALCRWGLLALLPFGAASSTOVCCTGTMKRLRSPETHLDMRLHYQCGVQVGNL 60
DB 1 MELALCRWGLLALLPFGAASSTOVCCTGTMKRLRSPETHLDMRLHYQCGVQVGNL 60
QY 61 ELTYLPTNASSPFIQDIOEVQGYVLINANQROYVLORLIRVGTQLFEDNVALAVLNG 120
DB 61 ELTYLPTNASSPFIQDIOEVQGYVLINANQROYVLORLIRVGTQLFEDNVALAVLNG 120
QY 121 DPLNNFNFTSFLRLRPKVSAS---HLEQLSLREILKGVLIORNPOLCVQDITLWK 176
DB 121 DPLNN-----TPPVGASGEGRIEQLSLREILKGVLIORNPOLCVQDITLWK 170
QY 177 DIFKNNQALTLIDITNRSRACHPCSPMKGSRCWGSSEDCQSLRTVCAGGACARCKGP 236
DB 171 DIFKNNQALTLIDITNRSRACHPCSPMKGSRCWGSSEDCQSLRTVCAGGACARCKGP 230
QY 237 LPTCCCHQCAAGCTGSKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTT 296
DB 231 LPTCCCHQCAAGCTGSKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTT 290
QY 297 FGASCVTACAPNYLSTVGSCTIVCPILHNOEYVAEDGTORCEKSPKCARVCYGLMEHL 356
DB 291 FGASCVTACAPNYLSTVGSCTIVCPILHNOEYVAEDGTORCEKSPKCARVCYGLMEHL 345
QY 357 REVAATVSAINQEPAGCKTIFGSLAFIPESFDDPMSNTAPLOP 400
DB 346 PPRPAVFPVLRMQPG--PAHPVLSFLRPSWDVSAFYSFLAP 387

```

RESULT 8
Q9UK79 PRELIMINARY; PRT; 419 AA.
Q9UK79;
AC Q9UK79;

RESULT 9
Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

RN (1)
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1; --
 KW Hypothetical Protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 1,1e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRFRTHQSDVSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGERLPQPICT 954
 DB 1 MALESILRRFRTHQSDVSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGERLPQPICT 60
 QY 955 IDVYIMVWCMMIDSECRPRFRELSEFSRMAPDQRFVVIQNEIDLPAASPLSTFYSRL 1014
 DB 61 IDVYIMVWCMMIDSECRPRFRELSEFSRMAPDQRFVVIQNEIDLPAASPLSTFYSRL 120
 QY 1015 LEDDMGDLVDAEYLVVQOGFCFDPAPAGAGVHHHRSSSTRSGGDLTLGLEPSEE 1074
 DB 121 LEDDMGDLVDAEYLVVQOGFCFDPAPAGAGVHHHRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHPSPLQRYSEDPYPLSEDTGY 1134
 DB 181 EPPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHPSPLQRYSEDPYPLSEDTGY 240
 QY 1135 APLTSPQPEYVNVQDVARPOPPSPREGPLPAAPAGATLERAKTLSPGKGVVDAFAFG 1194
 DB 241 APLTSPQPEYVNVQDVARPOPPSPREGPLPAAPAGATLERAKTLSPGKGVVDAFAFG 300
 QY 1195 GAVNEPEYLTQGGAAFPQPPPPAFSPAFDNLVYWDQPPRPGCAPSTFGTPTAENPEY 1254
 DB 301 GAVNEPEYLTQGGAAFPQPPPPAFSPAFDNLVYWDQPPRPGCAPSTFGTPTAENPEY 360
 QY 1255 LGLDVPV 1261
 DB 361 LGLDVPV 367

RESULT 10
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE POLYPROTEIN.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OC NCBI_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; --
 DR HSSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk PKinase.
 DR InterPro; IPR004028; Retro M.
 DR InterPro; IPR001245; Tyr PKinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02813; Retro M.; 1.
 DR ProDom; PD000001; Euk_Kinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFF1D53 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 8.5e-122;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTFCFGEAQCVCAGHKPPCCVARGSPGVPLDSYPMIMFPEDEGACQCPRI 634
 DB 141 PEETAPPTGP--DHCKCAHFIDGPHCVACCPGVGENDTL-VMKYADANAVCQLCHP 197
 QY 635 NCTHSCVDLDKGCFAEORASPLTSIVSAV--GILLVVLGVVFGILIKRQOKIRKRYTM 693
 DB 198 NCTRCCKPGLGCGP---NSGKTSIAAGVAGGLCLVVGGLGLRRR-HIVKRTL 253
 QY 694 RRLDQETLVLEPLPSGAMPQAOQRLTKETLRKVYVLSGAFYVYKQIMPDGENVK 753
 DB 254 RRLDQETLVLEPLPSGAMPQAOQRLTKETLRKVYVLSGAFYVYKQIMPDGENVK 313
 QY 754 IPVAIKVLENTSPKANKELIDEAYVAVAGVSPVSRLLGLCTSTVQLVTOIMPVGLL 813
 DB 314 IPVAIKELREATSRANKELIDEAYVAVADNPVNCGLLGLCTSTVQLVTOIMPVGLL 373
 QY 814 DHVENRGRLCSQDLVWCMQIAKMSYLEDVRLVHRDLAARVLYKSPHVKITDFGLA 873
 DB 374 DYIREHKDNISQYVLWVCVQIAKKNVYLEERLVRDLAARVLYKTPHVKITDFGLA 433
 QY 874 RLDDIDENYVADAGKVPKIMMALESILRRFRTHQSDVSYGVTTWELMTFGAKPYDGI 933
 DB 434 KLLGADSEYVYABGKPIKIMMALESILRHITVHQSDVSYGVTTWELMTFGSKPYDGI 493
 QY 934 AREIPDLLEKGERLPQPICTIDVYIMVWCMMIDSECRPRFRELSEFSRMAPDQRFV 993
 DB 494 ASEISSVLEKERLPQPICTIDVYIMVWCMMIDSECRPRFRELSEFSRMAPDQRFV 553
 QY 994 VTC-NEDLGPASPLDSTFYRSLLEDMDGDLVDAEYLVVQOGFCFDPAPAGAGVHHNR 105
 DB 554 VIQDERHNLPSPTDSKFYRLMEEDMEDIVDADETLVPHQGF----- 598
 QY 1053 HRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGASDVFDGDLGMAAGLQSL 110
 DB 599 -NSPST-----SRFTPLSSLSATSNNSATNCID-----RNGQGH 631
 QY 1108 PTHDPSPLQRYSEDPYPLPSET--DGVAPLTCSQPEYVNVQDVARPOPPSPREGPLPA 116
 DB 632 PVREDSEFYQRYSSDPTGNFLSESDIGFL-----PAPEYVNVQ--LMPKKPS----- 675
 QY 1166 ARPAGATLERAKTLSPGKGVVDAFV-----ARFGAVNEPEYL 1203
 DB 676 -----TAVYQNOIYNNISLTAISKLPMSRYQNSHSTAVDPEYL 715

RESULT 11
 Q86714 PRELIMINARY; PRT; 567 AA.
 AC Q86714;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE V-ERBB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OC NCBI_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities";

RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC6072.1; -.
 DR HSSP: P1362; IFGK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR01245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; TyrcK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 567 AA; 61390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 8,3e-122;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GREADQVACAHYKDPFCVACRPSGVKPDLSYMPFKPDEEGACOPCPINCHSCVDL 643
 DB 1 GP--DHCMKCAHFIDPCHVCACRPGVLENDTL-VKXADANAVCQLCPNCTRGCKGP 57
 QY 644 DDKGCPAQRASPLSTYSAYV-GILLVVIGVFGILIKRQCKIRKTYMRLLQETEL 702
 DB 58 GLEGCP---NGSKTBSIAGVVGGLCLVVGGLGILYLR--HIVRKRTLRLQEREL 113
 QY 703 VEFLTPSGAMPQACMRILKETELAKVVLGSGAGTYKGIWIPDGENYKIPAIYLR 762
 DB 114 VEPLTPSGEAPQAHRIKETEFKVKVLSGAGTYKGLWIPBEKVKIPAIKELR 173
 QY 763 ENTSPKANKELDEAYVAVAGVSPVSRLLGICLTSTVQLTQVLTQMPYGCILLDHVENRGR 822
 DB 174 EATSPKANKELDEAYVAVASVDNPRVCRLLGICLTSTVQLTQVLTQMPYGCILLDHVENRGR 233
 QY 823 LGSODLWMCQIACMGSTLEVDRLVHEDLAAKLVKSPHVKITDPGLARLLDIDETE 862
 DB 234 LOSQYLLWCVQIACMGSTLEVDRLVHEDLAAKLVKTPQHVKITDPGLAKLGADEKE 293
 QY 863 YHADGKVPKIMWALLESILRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942
 DB 294 YHAEKGKVPKIMWALLESILHRTYHQSDVMSYGVTVWELMTFGSKPFDGIPASISVLE 353
 QY 943 KGERLPQPICTIDVYMIMVCMYDSECRPRFELVSESRMARDPQRFVYIQ-NEDLG 1001
 DB 354 KGERLPQPICTIDVYMIMVCMYDADSRKFRFLIAEFSEKMARDPRLVYIQGDERMH 413
 QY 1002 PASPLDSTFYRSLDDDMGDLVDAEYLVPOQGFPCDPAPAGAGVWVHRRHSSSTRSG 1061
 DB 414 LPSPTDSKFRYRLMEEDMEDIVDADEVLPVHQGF-----NSPST--- 454
 QY 1062 GGDLLGLSEPEEERAPRSP-----APSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQ 1116
 DB 455 -----SRTPLLSLSTATSNNSATNCID-----RNGQGHPRREDSFPVQ 491
 QY 1117 RYSEDPVPLPSET--DGXVAPLTCSPQPEVYNQPDVRPQPSRREPLPAAPACATLE 1174
 DB 492 RYSSDPGNPLPESLDDGFL-----PAPEYVNO--IMPKKPS----- 526
 QY 1175 RAKTLPSPKNGVXKVF-----AFGAVENPEYL 1203
 DB 527 -----TAMVQNOIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL 566

RESULT 12
 Q8WYVO PRELIMINARY; PRT; 412 AA.
 AC Q8WYVO;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.

GN PP3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 growth.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318349; AAL55886.1; -.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR00719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; TyrcK; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.9e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 895 MALSEILRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 95
 DB 1 MALESILRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 60
 QY 955 IDVYIMVCMYDSECRPRFELVSESRMARDPQRFVYIQNEDLGASPLDSTFYRSL 10;
 DB 61 IDVYIMVCMYDSECRPRFELVSESRMARDPQRFVYIQNEDLGASPLDSTFYRSL 120
 QY 1015 LEDDMGGLVDAEYLVPOQGFPCDPAPAGAGVWVHRRHSSSTRSGGDLTLGLEPSEE 10
 DB 121 LEDDMGGLVDAEYLVPOQGFPCDPAPAGAGVWVHRRHSSSTRSGGDLTLGLEPSEE 186
 QY 1075 EAPRSPLAPSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQRYSEDPVPLPSETGGYV 113
 DB 181 EAPRSPLAPSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQRYSEDPVPLPSETGGYV 240
 QY 1135 APLTCSPPQPEVYNQPDVRPQPSRREGLPAAPACATLEPAKTLSPKNGVXKVFVAFG 119
 DB 241 APLTCSPPQPEVYNQPDVRPQPSRREGLPAAPACATLEPAKTLSPKNGVXKVFVAFG 300
 QY 1195 GAVENPEYLTQGGAAAPQ-----HPPA---FSPAFDNL 122
 DB 301 GAVENPEYLTQGGAAALSPILLPSAPSTTITGTRTHQSGGHPAPSGHURORTOST 360
 QY 1227 YYWD-QDPPER-----GAPSTFGKGTPTAEN 1251
 DB 361 WWTQCEPEEGVYRRSPDVSSGSHGLTSAGIKRWEGPPTTSGTCAHY 410

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-v-erb-A-v-erb-B protein.
 GN Gag-v-erb-A-v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxId=11861;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=90206603; PubMed=1969616;
RA Bruckin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B";
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA36459.1; -
DR EMBL; X52211; CA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_1ig.
DR InterPro; IPR001723; Strchnm_receptot.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Csteroi.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zfc4; 1.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00447; STROIDPINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_Csteroi; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5MED791E895CE CRC64;

Query Match 24.1%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1,4e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 547 VEEGRVQLPRE-YVNAR-HCLP-----CHPEQC 574
DB 354 IEKQESYLAFEHYIYRKHNIHFMSKLMKVADLRMGAYHASFILMKVECPTELS 413
QY 575 PONGSVTCFGEADQCAHYKDPFCVACPSGVKPDLSYMPKPFDEGACOPCPPI 634
DB 414 PGE-----VGP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAQCQLCHP 465
QY 635 NCTHSCVLDLDDKGPACQASPLTYSVANV-GILLVVVLGVVFGILLRRQOKIRKXTM 693
DB 466 NCTRGCKGPGLEGP--NGSKTPSIAAGVAGGLCLVYVVGIGIGLYLRR-HIVKRTL 521
QY 694 RRLQETELVEPTPSGAMPNQAMRIKETELRKVKVLSGAFGVYVGIWIPGSENVK 753
DB 522 RRLQERELVEPLPSEARNQAHRLIKETEFKKVKVLFQFAGFVYVKGILPGEKVT 581
QY 754 IPAVAKVLRNTPSKANKETLDEAYVMAGVSPYVSRLLIGICTSTVQLTOLMPEYGL 813
DB 582 IPAVAKELREATSPKANKETLDEAYVMASVDNPHVCRLLIGICTSTVQLITQLMPEYGL 641
QY 814 DHYRENRGRGSDLLNMCMQIAKMSYLEDRVLYHRDLAARNVLYKSNHYKITDEGLA 873
DB 642 DYIREHKDNIGSYLLNMCVQIAKGMVYLEBRNHRDLAARNVLYKTPQHYKITDFGLA 701
QY 874 RLIDIDETEHADGKVPKIMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGP 933
DB 702 KQLGADKEHYHAGGVPIKMMALLESILHRYTHQSDVMSYGVTVWELMTFGSKRYDIP 761
QY 934 AREIPULEKGERLPPPICTIDVYVIMYKCMIMSECCPRRELVSFPMARPPQGFV 993
DB 762 ASSTSVLEKGERLPPPICTIDVYVIMYKCMIMSGDSRPKRELIAFSKQARPPREYL 821
QY 994 VIQ-NEDLGPAFLDSTFYRSLLEDQMDGLVDAEYVLPQGFPCPDPAAGAGMVMHR 1052
DB 822 VIQDERMHLPSFTDSKPYRTIMEEDVEDIVADDEYLVPHGGF----- 866

QY 1053 HRSSSTRGGDILLGEPSEEARPSRIAPASEAGSDVFDGLGMGAAGLQSLPRTHP 111
DB 867 -NSPT-----SRTPLLSSLSATSN-----NSATCICIRNGCH-- 896
QY 1113 SPLQRYSEDPVPPLPSETDGYAPRLTCSPOPEYVNCQDVPPOPSPREGPLPAAPAGAT 117
DB 899 -----PVNEDGL-----PAPEYVNC--LMPKAPSTAMONQIYNYISLT 936
QY 1173 -LERAKTLSPKNGVNVKQVPAFGAVENPEYL 1203
DB 937 AISKLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
ID 085468 PRELIMINARY; PRT; 545 AA.
AC 085468;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (T834) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA "Scotling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA30024.1; -
DR HSSP; P13362; 1RGX.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase;
KW Transcription regulation; Tyrosine-protein kinase;
SQ SEQUENCE 545 AA; 60899 MW; 140DC8E8CA0F8AF4 CRC64;

Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.7e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

QY 584 GPEADQVACAHYKDPFCVACPSGVKPDLSYMPKPFDEGACOPCPINCTHSCVDL 643
DB 1 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAQCQLCHPCTGCKGSP 57
QY 644 DDKGPAEGRASPLTYSVANV-GILLVVVLGVVFGILLRRQOKIRKXTMRLLQETEL 702
DB 58 GLEGP--NGSKTPSIAAGVAGGLCLVYVVGIGIGLYLRR-HIVKRTLRLQEREL 113
QY 703 VEPITPSGAMPNQAMRIKETELRKVKVLSGAFGVYVGIWIPDGENVKIPIVAKYL 762
DB 114 VEPITPSGAPNPQAHRLIKETEFKKVKVLFQFAGFVYVKGILPGEKVTIPIVAKELR 173
QY 763 ENTPSKANKETLDEAYVMAGVSPYVSRLLIGICTSTVQLTOLMPEYGLDHYRENRGR 822
DB 174 EATSPKANKETLDEAYVMASVDNPHVCRLLIGICTSTVQLITQLMPEYGLDYIREHKDN 233
QY 823 LGSQDILNMCMQIAKMSYLEDRVLYHRDLAARNVLYKSNHYKITDGLRLIDDETE 882
DB 234 IGSQYLLNMCVQIAKGMVYLEBRNHLVHRDLAARNVLYKTPQHYKITDFGLAKQAGADKE 293
QY 883 YHADGKVPKIMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGPAREIPDLE 942
DB 294 YHAEQKVPKIMMALESILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSYLE 353

```

QY 943 KGERLPDPICITIDVYIMVCMKMSIDSECRPREPLVSEFSEFARMARDQRFVIO-NEDLG 1001
DB 354 KGERLPDPICITIDVYIMVCMKMSIDSECRPREPLVSEFSEFARMARDQRFVIO-NEDLG 413
QY 1002 PASPLDSTFFRSLBDDMDGLVDAEYLVQOGFCFCDPAAGAGVHHRRSSSTRSG 1061
DB 414 LPSPTDSKFYRTLMSEEDMEDIVDAEYLVPHQGF-----NSPST---454
QY 1062 GGDULTGLSEPESEFARSPRL-----APEGAGSDVFDGDLGMAKGLQSLPHIDPSPLQ 1116
DB 455 -----SRPLSSLSATSNMNTATNCIDRNGS-----H-----481
QY 1117 RYSEDPVPLPSEITDGYAPLPTCSQPEYVQNPDPVPPSPREGLPAAPAGAT-LER 1175
DB 482 -----PYREDFGL-----PAPRYVQ--LMPKPESTAVQIQINYSLSLTKSK 523
QY 1176 AKTLSPGKGVKXDFAFGAGAVENEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNEYL 544

RESULT 15
Q9WVF5 PRELIMINARY: PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RA STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibhe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR genes Encode Carboxy-Terminal Truncated Receptors.";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N. A.
RA STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampard A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N. A.
RA STRAIN=C57BL/6J; TISSUE=LIVER;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pedole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner V., Washio T.,
RA Sakai K., Okido T., Furuno M., Kono H., Badarrelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL: AF124513; AAD4149.1; -
DR EMBL: AF275366; AAG38047.1; -
DR EMBL: AF275366; AAG38047.1; JOINED.
DR EMBL: AF275365; AAG38047.1; JOINED.
DR EMBL: AK004944; BAB23688.1; -
DR EMBL: AK004883; BAB23641.1; -
DR EMBL: AK004911; BAB23662.1; -
DR MGD: MGI:95294; Egfr.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like.
DR Pfam: PF01030; Recep_L domain; 2.
DR SMART: SMO0261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 21.9%; Score 1497.5; DB 11; Length 655;
Best Local Similarity 43.3%; Pred. No. 5,7e-105;
Matches 280; Conservative 100; Mismatches 214; Indels 33; Gaps 8

```

```

QY 11 LLLALLPFGA--STOYCTGDMKLRLPAPETHLMDLRLYGGCGVQGNLETLPLPT 68
DB 14 LLLALLPFGA--STOYCTGDMKLRLPAPETHLMDLRLYGGCGVQGNLETLPLPT 73
QY 69 ASLFLDIOGVGVYLIHNOVQVQLRRLVGTQTFEDVYALAVLDNGPLNNFN 128
DB 74 YDLSPKTIQVAVYLIHNOVQVQLRRLVGTQTFEDVYALAVLDNGPLNNFN 126
QY 129 FTVSFWLRVPVSAHSHQ--SLTILKGVLIQNPOLCTQDTILMKDT----FKN 182
DB 127 -----TNRGTRELPMRLQILLGAARFSNNPLTCMDITQWIDVQNVFMSN 175
QY 183 NQALTLIDITRSRACHPCSPMCKSRGCESESDCSLRTVACAGCA-RCKGPLPTDC 241
DB 176 MSMDL-----QHPSSCKCKDPSGNSGCKGCEGNOCKLTKIICAGQCSKRCGRSPDC 221
QY 242 CHEQCAAGCTGPKASDCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 301
DB 232 CHNOCAAGCTGPKASDCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 291
QY 302 VTACPVYVLTSDVSCCTLVCPRLHNOEVTAEFGORCKSKPCARVCYGLMEHLEVA 361
DB 292 VKKCPRIYVYVTDHSCVYRACGPYIEV-EDGIRKCKCKDGPCKVCKNGIGEPYDTLS 350
QY 362 VTSANIGFAGCKKIFGSLAFLPESFDDGASNTAPLOPEQLQVFTLEITGYLYISAM 421
DB 351 INATNIGFAGCKKIFGSLAFLPESFDDGASNTAPLOPEQLQVFTLEITGYLYISAM 410
QY 422 PDSLPDLSVQNLQVIRGRILNHGASVLTGIGISMLGIRSRRELGSGLALHNHTHC 481
DB 411 PDNWTDLHAFENLEIRGRIRKQHGQSLAVGANTISLGRSKETSDVDVITISGRNLC 470
QY 482 FVATVPMDQLFRNHQALHTANRPEDEGEGLACHQICARHCWGPGPTCVNCSQL 541
DB 471 YANTINIKKLFGRNQTAKIMNRAEDCAVNVHNPCLSSGCGCGPPROCVSQQNVS 530
QY 542 RGQSCVECEKVLQGLPREYVNAHCLCPHCECPQNGSVTTCGPEZDQVACAHYKDPPE 601
DB 531 RGRCEVCKVITLGEPEFENSECICQHECLPQMAVNTITCTGRGDNICQCAHYIDGPH 590
QY 602 CVARCPGVKPDLSYPIWKFPEEGACPCPINCCHSCVTDLDDKGC 648
DB 591 CVKCPAGINGENNTL-VKTYALANNVCHLCHANCYTGAGAGELQGC 636

```

Search completed: July 22, 2003, 09:00:23
Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds

(without alignments)

4891.279 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 6815

Sequence: 1 MELALCRWGLLALLPFGA.....TFKGTPTAEKPEYLGLDVPY 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6640	97.4	1255	1	ERB2_HUMAN
2	5849	85.8	1257	1	ERB2_RAT
3	5833.5	85.6	1254	1	ERB2_MESAU
4	3065	45.0	1210	1	EGFR_HUMAN
5	3039	44.6	1210	1	EGFR_MOUSE
6	2887.5	42.4	1308	1	ERB4_HUMAN
7	2868	42.1	1308	1	ERB4_RAT
8	2614.5	38.4	1167	1	XMRK_XIPMA
9	2341.5	34.4	1342	1	ERB3_HUMAN
10	2269.5	33.3	1339	1	ERB3_RAT
11	1897	27.7	1426	1	EGFR_DROME
12	1669.5	24.5	634	1	ERBB_ALY
13	1623	23.8	604	1	ERBB_AVIER
14	1595	23.4	703	1	EGFR_CHICK
15	1558	22.9	540	1	ERBB_AVEU
16	1243	18.2	1332	1	LT22_CABEL
17	1142.5	16.8	1243	1	ERB2_MOUSE
18	700	10.3	1363	1	ILPR_BRALA
19	666	9.8	1383	1	INSR_RAT
20	665.5	9.8	1372	1	INSR_MOUSE
21	663	9.7	1300	1	IRR_MOUSE
22	662	9.7	1382	1	INSR_HUMAN
23	658	9.7	1477	1	HTK_HYDAT
24	657.5	9.6	1300	1	IRR_CAVPO
25	655	9.6	1607	1	MIFR_LYMT
26	654	9.6	1297	1	IRR_HUMAN
27	602	8.8	1367	1	IGIR_HUMAN
28	591	8.7	1390	1	INSR_AEDAE
29	590	8.7	2146	1	INSR_DROME
30	584	8.6	1373	1	IGIR_MOUSE
31	580.5	8.5	1370	1	IGIR_RAT
32	572	8.4	987	1	EPB4_HUMAN
33	565	8.3	1114	1	RET_HUMAN

34	563.5	8.3	984	1	EPB1_CHICK	C07494	gallus gall
35	562.5	8.3	977	1	EPB2_MOUSE	O03145	mus musculu
36	560.5	8.2	984	1	EPB1_RAT	P09759	rattus norv
37	554.5	8.1	984	1	EPB1_HUMAN	P54762	homo sapien
38	552.5	8.1	976	1	EPB2_HUMAN	P29317	homo sapien
39	552	8.1	902	1	EPBB_XENLA	O91736	xenopus lae
40	552	8.1	987	1	EPB4_MOUSE	P54761	mus musculu
41	544.5	8.0	985	1	EPBA_XENLA	O91571	xenopus lae
42	544	8.0	757	1	HT16_HYDAT	P53356	hydra atten
43	544	8.0	1068	1	PAK1_XENLA	O91738	xenopus lae
44	542.5	8.0	1053	1	PAK1_CHICK	O00944	gallus gall
45	532.5	7.8	1052	1	PAK1_MOUSE	P34152	mus musculu

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MDN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.,
RA Cossens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seedurg P.H., Liberman T.A., Schlesinger J.,
RA Francis U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN (3)
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN (4)
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
RN (5)
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY)
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M1767; AAA35808.1; -
 DR EMBL: M1761; AAA35808.1; JOINED.
 DR EMBL: M1762; AAA35808.1; JOINED.
 DR EMBL: M1763; AAA35808.1; JOINED.
 DR EMBL: M1764; AAA35808.1; JOINED.
 DR EMBL: M1765; AAA35808.1; JOINED.
 DR EMBL: M1766; AAA35808.1; JOINED.
 DR EMBL: M1730; AAA5493.1; -
 DR EMBL: M12036; AAA5978.1; -
 DR EMBL: X03363; CAA27060.1; -
 DR PIR: A25491; A25491.
 DR PIR: A24571; A24571.
 DR HSSP: P1362; 1FGK.
 DR Genew: HGNC:3430; ERBB2.
 DR MIM: 164870; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	643	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTID=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	P -> A (IN REF. 2).
FT	CONFLICT	1255	1255	Best Local Similarity 97.5%; Pred. No. 0; Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;
SEQ	SEQUENCE	1255	AA: 137909	MM: 3983DFDA04DC962

Query Match 97.4%; Score 6640; DB 1; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY	1	MELALCRWGILLALPPGAASVCTGCTDMKRLPASPEHLDMLRHLYOGCCVVGNL	60
DB	1	MELALCRWGILLALPPGAASVCTGCTDMKRLPASPEHLDMLRHLYOGCCVVGNL	60
QY	61	ELTYLPNTASISFLDIDIEVGAYLIHNOVRVPLRLRVGTOLFEENYALAVDNG	120
DB	61	ELTYLPNTASISFLDIDIEVGAYLIHNOVRVPLRLRVGTOLFEENYALAVDNG	120
QY	121	DLNNTTVTASPGGRLRLSLTEILKGVILIQNPOLCYQDTILMKDIFKNNOLA	180
DB	121	DLNNTTVTASPGGRLRLSLTEILKGVILIQNPOLCYQDTILMKDIFKNNOLA	180
QY	181	LTLLDITNRSRACHPCSPCKSGSRCKMGSSSDCOSLRTVAGGACARCKGLPDDCCEQC	240
DB	181	LTLLDITNRSRACHPCSPCKSGSRCKMGSSSDCOSLRTVAGGACARCKGLPDDCCEQC	240
QY	241	AAAGCTPGHSDCLALPHNHSIGICELCPALVTYNTDFESMENPEGRYTFGASCTYACF	300
DB	241	AAAGCTPGHSDCLALPHNHSIGICELCPALVTYNTDFESMENPEGRYTFGASCTYACF	300
QY	301	YNYLSTDVSGCTIYCPILNQEVYTHEDGTQCEKSKPCAVCYGLMGQYTKANSKFTIGIT	360
DB	301	YNYLSTDVSGCTIYCPILNQEVYTHEDGTQCEKSKPCAVCYGLMGQYTKANSKFTIGIT	360
QY	361	ELERAGCKTIFGSLAFPLPESFDGDPASNTAPLOPEQLQVETLEETGYLYISAWPDSLP	420
DB	361	ELERAGCKTIFGSLAFPLPESFDGDPASNTAPLOPEQLQVETLEETGYLYISAWPDSLP	420
QY	421	DLSPQNLQVIRGILINNGAYSLTGLGISTGLSLRELSSGLALIHNNTHLCFVHTV	480
DB	421	DLSPQNLQVIRGILINNGAYSLTGLGISTGLSLRELSSGLALIHNNTHLCFVHTV	480
QY	481	PMDDLFNPPHQAALHTANRPEDEVGAGLACHOLCARGHCMGPGPOCNVCSQFLMGQC	540
DB	481	PMDDLFNPPHQAALHTANRPEDEVGAGLACHOLCARGHCMGPGPOCNVCSQFLMGQC	540
QY	541	VEECRVYQGLPREVYVNRHCLPCHPECPQNGSVTCFGEDADQCVACAHYKDPFCVAC	600
DB	541	VEECRVYQGLPREVYVNRHCLPCHPECPQNGSVTCFGEDADQCVACAHYKDPFCVAC	600
QY	601	PSGVKPLSTMPIMKPFDEBGAQCPINCTHSCVDLDKRGCPAEOBPASTLSISAVVG	660
DB	601	PSGVKPLSTMPIMKPFDEBGAQCPINCTHSCVDLDKRGCPAEOBPASTLSISAVVG	660
QY	661	ILVVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBVLTPSGAMPNQAQRILKEQYI	720
DB	661	ILVVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBVLTPSGAMPNQAQRILKEQYI	720


```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138631 MW; 6129264583011402 CRC64;

```

Query Match 85.8%; Score 5849; DB 1; Length 1257;
 Best Local Similarity 85.9%; Pred. No. 4,2e-309;
 Matches 1081; Conservative 55; Mismatches 118; Indels 4; Gaps 4;

```

QY 1 MELALCRWGILLALLPPGAASVQVCTGTDKRLPASPETHLDMLRHLYOGGVQGNL 60
DB 1 MELAMCMWGLLALLPPGILACTGCTGDKLPPASPTHLDMLRHLYOGGVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIANQYQVPLQRLRIVRGTOLEFEDNYALVLDNG 120
DB 61 ELTYVPANASLFLQDIOEVQGYVLIANQYKRVPLQRLRIVRGTOLEFEDKXVALVLDNR 120
QY 121 DELNNTTPTV-GASGGIREQLRLSTLEILKGVLIQRNPQLCYQDTILMKD.FHKNQNL 179
DB 121 PDQNVAASTPGRTPEGLRELQRLSTLEILKGVLIQRNPQLCYQDVKLVKDFVRKNQNL 180
QY 180 ALLLIDTNRSPACHPCSPMCKSGRQWGESSEDCSLTRTYCAGGACARCKGLPTDCCHEO 239
DB 181 APVDIDTNRSPACHPCAPACKDNHGWGESPEDQILGTICTSGACARCKRLPTDCCHEO 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTESPNRBEGRYTFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTESPNRBEGRYTFGASCVTAC 300
QY 300 PNYLSTVSGSCTLVCPHNOEVAEDGTORCEKCSFPCARVCYGMQYIKANSKFI 359
DB 301 PNYLSTVSGSCTLVCPHNOEVAEDGTORCEKCSFPCARVCYGMQYIKANSKFI 360
QY 360 TELEPAGCKKIFGSLAPLPSFDDDPASNTAPLQPEQLQVFETLEITGYLISAMPDSL 419
DB 361 NVQSFDDGCKKIFGSLAPLPSFDDDPASNTAPLQPEQLQVFETLEITGYLISAMPDSL 420
QY 420 PDLVFNOLQVIRGRILLNGAYSLTLQGLISWGLRSARELSSGALLHNTHLCFVHT 479
DB 421 RDLVFNOLQVIRGRILLNGAYSLTLQGLISWGLRSARELSSGALLHNTHLCFVHT 480
QY 480 VPMQDLFRNPHQALLHTANREDE-CVGEGLACHOLCARCHGWPGPTQVNCOSQFLRQ 538
DB 481 VPMQDLFRNPHQALLHTANREDE-CVGEGLACHOLCARCHGWPGPTQVNCOSQFLRQ 540
QY 539 ECVEECRVLOGLPREYNNARHCLPCHECOPQNGSVTCFEPAPDQVAAAHYKDPFCYA 598
DB 541 ECVEECRVKGLPREYVSDKRCPCHECCPQNSSEFCFSESDQCAAAHYKSSSCYA 600
QY 599 RCPGKVPDLSYMPIMKFPDEGACQPCPINCCHSCVLDLDCKCPAPRASPLTSIVASV 658
DB 601 RCPGKVPDLSYMPIMKFPDEGACQPCPINCCHSCVLDLDCKCPAPRASPLTSIVASV 660
QY 659 VGIILVVVLGVVFIILKRCQOKIKRYTWRLLQFELVEPLTPSGAMPNQOAMRLKEQ 718
DB 661 VGIILVVVLGVVFIILKRRRQOKIKRYTWRLLQFELVEPLTPSGAMPNQOAMRLKET 720
QY 719 YIKANSKFIGITEL-TVYKGIWIPDGENVKIPVAIVLENTSPKANKRILDEAVYMGV 777
DB 721 ELR-KYVVLGSGAGTYKGIWIPDGENVKIPVAIVLENTSPKANKRILDEAVYMGV 779
QY 778 GSPYVSRLLGICLTSTVQVLTQMLPFQCLLDHYRENRLGSGODLLNMCQIAKGSYLE 837

```

```

DB 780 GSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYREHNRGLSGODLLNMCQIAKGSYLE 839
QY 838 DVRLVHRDLARNAVLYVSPNVRKITTDFGLARLLDIDETEHADGKVPKIMMALESILRR 897
DB 840 DVRLVHRDLARNAVLYVSPNVRKITTDFGLARLLDIDETEHADGKVPKIMMALESILRR 899
QY 898 RFTYQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMYK 957
DB 900 RFTYQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMYK 959
QY 958 CMMTDSERCPRETLVEFSRMAPDRPFVIONEDLGPASPLDSFFYSLDEDDMDGL 1017
DB 960 CMMTDSERCPRETLVEFSRMAPDRPFVIONEDLGPASPLDSFFYSLDEDDMDGL 1019
QY 1018 VDABEYLVPOQGFPCPPAPAGQGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSPLAP 1077
DB 1020 VDABEYLVPOQGFSPDPTOTGSTARRRRSSSTRSGGDLTLGLEPSEEGPSPPLAP 1079
QY 1078 SEGAGSDVFDGDLGMGAAGLQSLPTHDEPFLQRYSEDPVPLPSETDGVAPLTCSPQ 1137
DB 1080 SEGAGSDVFDGDLGMGYTKGLQSLPHDLSPQRYSEDPVPLPSETDGVAPLTCSPQ 1139
QY 1138 EYVNAQPDVRRPQPSRGPPLPAPRAGATLERAKTISPCKAGVYKDVAFGAVENPEYL 1197
DB 1140 EYVNAQSDVQPPPLTPGCPPLPVPBPATLERKITSFGKNGVYKDVAFGAVENPEYL 1199
QY 1198 TPQGAAPQPPPPAPSPAFDNTLYWQDQPERGAPSPFTFKGTAEENEYGLDVPV 1255
DB 1200 VPREGTASPPHSPAFSPAFDNTLYWQDQSEGGPPSPNEGTAEENEYGLDVPV 1257

RESULT 3
ERR2_MESAU STANDARD; PRT; 1254 AA.
ID ERR2_MESAU
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DR 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor Protein-tyrosine kinase etdb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
KA [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.

DR HMBL: D16295; BAC03801.1; -
 DR HMBL: P11367; JFCK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 22 1254
 FT TRANSMEM 652 652
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 584 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT CARBOHYD 658 658
 FT VARIANT 659 659
 FT VARIANT 659 659
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C31F2BE1 C6C64;

Query Match 85.6%; Score 5833.5; DB 1; Length 1254;
 Best Local Similarity 85.5%; Pred. No. 2.9e-308;
 Matches 1074; Conservative 64; Mismatches 115; Indels 3; Gaps 3;

QY 1 MELALCRWGLLLALPPGASTOVCTGTDMLRLPASPETHLMDLRHLYOGGQVVGNTL 60
 Db 1 MELAAWGMGLLLALLSPGASGTQVCTGTDMLRLPASPETHLMDLRHLYOGGQVVGNTL 60

QY 61 ELTYLPTNASLSPFDIOIQGVGVLIANOVROVPLQRLIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPANNTLSFLDDIOEVGVYMLIAHSOVHPLQLRLIVRGTLQFEDNYALAVLDNR 120
 QY 121 DPLNNTPTVPGASPGGLRELOQRSLETLKGVLIQRNPOLCYOPTIMKDIPIKNNOLA 180
 Db 121 DPLDNTTATGTATPGGLRELOQRSLETLKGVLIQRNPOLCYOPTIMKDIPIKNNOLA 180
 QY 181 LTLIDTNRSPACHPSPMCKSGRCWSESSDCCSLTRTVACAGCARCKGRLPTDCHEQC 240
 Db 181 PYDIDTNRSPACHPSPMCKSGRCWSESSDCCSLTRTVACAGCARCKGRLPTDCHEQC 240
 QY 241 AAGCTGPKRSDCLACHTNHSICILHCPALVTYNTDFFESMPREGYRTGACVYACP 300
 Db 241 AAGCTGPKRSDCLACHTNHSICILHCPALVTYNTDFFESMPREGYRTGACVYACP 300
 QY 301 YNYLSTDVGSCTLVCPPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTEVGSCTLVCPPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMHLAGALITSAN 360
 QY 361 ELFFAGCKKIFGSLAFPESTDQDPASNTAPLOEBOLOVFETLEITGYLISAMPDLP 420
 Db 361 IQEFAGCKKIFGSLAFPESTDQDPASNTAPLOEBOLOVFETLEITGYLISAMPDLP 420
 QY 421 DLSVFNQNLVIRGRILHNGAYSILTQGLGIGLGRSLRELSGLALIHNTHTLCFVHTV 480
 Db 421 DLSVFNQNLVIRGRILHNGAYSILTQGLGIGLGRSLRELSGLALIHNTHTLCFVHTV 480
 QY 481 PWDQFLPNPHOALHTNANPEDECGVEGLACHQLCARHCWCPGPTCVNCSQFLRQEC 540
 Db 481 PWDQFLPNPHOALHTNANPEDECGVEGLACHQLCARHCWCPGPTCVNCSQFLRQEC 540
 QY 541 VEEGVNLGLREVYNANHCPCPEQCPQNGASTCCGPEBDQCVAAHYKDPFCVARC 600
 Db 541 VEEGVNLGLREVYNANHCPCPEQCPQNGASTCCGPEBDQCVAAHYKDPFCVARC 600
 QY 601 PSQVFPDLSYMPIMKFPDEBACQPCPCINCTHSCVDLDDGCPAEOGASPLTSIVAVG 660
 Db 601 PSQVFPDLSYMPIMKFPDEBACQPCPCINCTHSCVDLDDGCPAEOGASPLTSIVAVG 660
 QY 661 ILVVVGLGVGILIKRQOKIRKTYMRLLQETLEVEPLTPSGAMPNQAMRLKEQYI 720
 Db 661 ILVVVGLGVGILIKRQOKIRKTYMRLLQETLEVEPLTPSGAMPNQAMRLKEQYI 720
 QY 721 KANSKFIGITELTVYKGIWIPDGENVKIPAIKIVRENTSPRANKELIDEAVVMGVGS 779
 Db 721 R-KYKVLSSGAFGVYVIGIWIPOGENVKIPAIKIVRENTSPRANKELIDEAVVMGVGS 779
 QY 780 PYVBRLLGICLTSTVQVLTQMPYGCILDHVRENRGLSGQDILLNMCQIAKMSYLEDV 839
 Db 780 PYVBRLLGICLTSTVQVLTQMPYGCILDHVRENRGLSGQDILLNMCQIAKMSYLEDV 839
 QY 840 RLYVRDLAARVAVLKSNNHKTDFGLARLLDIDETEHYADGKVPKMALESILRRF 899
 Db 840 RLYVRDLAARVAVLKSNNHKTDFGLARLLDIDETEHYADGKVPKMALESILRRF 899
 QY 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPASEIPDLLEKGERLPQPICTIDVYMIWVCM 959
 Db 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPASEIPDLLEKGERLPQPICTIDVYMIWVCM 959
 QY 960 MIDSECPREELVSESSRNARPPQRPVYIQNEDLPASFLDSTFYRSLLEDDMDGLVD 1019
 Db 960 MIDSECPREELVSESSRNARPPQRPVYIQNEDLPASFLDSTFYRSLLEDDMDGLVD 1019
 QY 1020 ABEYLVPOQGFCEPDPAAGAWHRRHSSSTRSGGDLTGLFSESEEARSPAPSE 1079
 Db 1020 ABEYLVPOQGFCEPDPAAGAWHRRHSSSTRSGGDLTGLFSESEEARSPAPSE 1079
 QY 1080 GAGSDVVDQDLGMAAKGLQSLFTHDPSPLQRYSEDPTVPLSEITDGYVAPVLTGSPQEPY 1139
 Db 1080 GAGSDVVDQDLGMAAKGLQSLFTHDPSPLQRYSEDPTVPLSEITDGYVAPVLTGSPQEPY 1139

```

QY      1140 VNQDVRPQPPSPRGDPLPAAPACATIERAKTISPGKNGVYKQVFAFGAVENPEVLT 1199
DB      1140 VNQPEVRPQPLTPGSPPLPPAPACATIERPKTISPGKNGVYKQVFTFGAVENPEVLT 1199
QY      1200 OGGAAPQHPHPAPAFPAFDNLYYMDQDPPERCAPSTFKGTPTAENPEYGLADVPV 1255
DB      1200 RGGSASQPH-PPALCPRAFDNLYYMDQDPSERKSPNTTEGFTAENPEYGLADVPV 1254

RESULT 4
EGFR_HUMAN          STANDARD:      PRT: 1210 AA.
ID      P00533; P06268; Q14225; Q9UMD8; Q9UMG5; Q92735; C00732;
AC      C00698; Q9BES2; Q9H2C9; Q9GZX1; Q9H3C9;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
GN      protein-tyrosine kinase ErbB-1).
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RX      MEDLINE=84219729; PubMed=6328312;
RA      Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA      Lee Y., Yarden Y., Libermann T.A., Schlessinger U., Downward J.,
RA      Mayes E.L.V., Whittle N., Waterfield M.D., Seedbarg P.H.;
RT      "Human epidermal growth factor receptor cDNA sequence and aberrant
RT      expression of the amplified gene in A431 epidermoid carcinoma cells.";
RT      Nature 309:418-425 (1984).
RN      [2]
RX      MEDLINE=95382957; PubMed=7654368;
RA      Ilekis J.V., Stark B.C., Scoccia B.;
RT      "Possible role of variant RNA transcripts in the regulation of
RT      epidermal growth factor receptor expression in human placenta.";
RT      Mol. Reprod. Dev. 41:149-156 (1995).
RN      [3]
RX      MEDLINE=97078686; PubMed=8918811;
RA      Reiter J.L., Mahlie N.U.;
RT      "A 1.8 kb alternative transcript from the human epidermal growth
RT      factor receptor gene encodes a truncated form of the receptor.";
RT      Nucleic Acids Res. 24:4050-4056 (1996).
RN      [4]
RX      MEDLINE=97256547; PubMed=9103388;
RA      Ilekis J.V., Garliti J., Niederberger C., Scoccia B.;
RT      "Expression of a truncated epidermal growth factor receptor-like
RT      protein (TEGFR) in ovarian cancer.";
RT      Gynecol. Oncol. 65:36-41 (1997).
RN      [5]
RX      MEDLINE=21100872; PubMed=1161793;
RA      Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA      Scheil Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampfand A.L.,
RA      Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.;
RT      "Comparative genomic sequence analysis and isolation of human and
RT      mouse alternative EGFR transcripts encoding truncated receptor
RT      isoforms.";
RT      Genomics 71:1-20 (2001).
RN      [6]
RX      MEDLINE=84219729; PubMed=6328312;
RA      Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA      Lee Y., Yarden Y., Libermann T.A., Schlessinger U., Downward J.,
RA      Mayes E.L.V., Whittle N., Waterfield M.D., Seedbarg P.H.;
RT      "Human epidermal growth factor receptor cDNA sequence and aberrant
RT      expression of the amplified gene in A431 epidermoid carcinoma cells.";
RT      Nature 309:418-425 (1984).
RN      [7]
RX      MEDLINE=84196372; PubMed=6326261;
RA      Lin C.R., Chen W.S., Kruller W., Stolarek L.S., Weber W.,
RA      Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT      "Expression cloning of human EGF receptor complementary DNA: gene
RT      amplification and three related messenger RNA products in A431
RT      cells.";
RT      Science 224:843-848 (1984).
RN      [8]
RX      MEDLINE=84245835; PubMed=6330563;
RA      Xu Y.H., Ishii S., Clark A.J., Sullivan M., Wilson R.K., Ma D.P.,
RA      Roe B.A., Merlino G.T., Pastan I.;
RT      "Human epidermal growth factor receptor cDNA is homologous to a
RT      variety of RNAs overproduced in A431 carcinoma cells.";
RT      Nature 309:806-810 (1984).
RN      [9]
RX      MEDLINE=85046483; PubMed=6093780;
RA      Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA      O'Malley B.W.;
RT      "Isolation of an evolutionarily conserved epidermal growth factor
RT      receptor cDNA from human A431 carcinoma cells.";
RT      Biochem. Biophys. Res. Commun. 124:125-132 (1984).
RN      [10]
RX      MEDLINE=88217333; PubMed=3329716;
RA      Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA      Waterfield M.D.;
RT      "The human EGF receptor gene: structure of the 110 kb locus and
RT      identification of sequences regulating its transcription.";
RT      Oncogene Res. 1:375-396 (1987).
RN      [11]
RX      MEDLINE=91107677; PubMed=1988448;
RA      Haley J.D., Waterfield M.D.;
RT      "Contributory effects of de novo transcription and premature
RT      transcript termination in the regulation of human epidermal growth
RT      factor receptor proto-oncogene RNA synthesis.";
RT      J. Biol. Chem. 266:1746-1753 (1991).
RN      [12]
RX      MEDLINE=85270438; PubMed=2991899;
RA      Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT      "Characterization and sequence of the promoter region of the human
RT      epidermal growth factor receptor gene.";
RT      Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924 (1985).
RN      [13]
RX      MEDLINE=89278137; PubMed=2543678;
RA      Margolis B.L., Lax I., Kris R., Domagala M., Horigger A.M.,
RA      Howk R., Givol D., Ullrich A., Schlessinger U.;
RT      "All autophosphorylation sites of epidermal growth factor (EGF)
RT      receptor and HER2/neu are located in their carboxyl-terminal tails.
RT      Identification of a novel site in EGF receptor.";
RT      J. Biol. Chem. 264:10667-10671 (1989).
RN      [16]

```

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 the epidermal growth factor receptor expressed in Chinese hamster
 ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 Asn-X-Cys sequence of recombinant human epidermal growth factor
 receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 truncated isoform/TEGFR, 3/p110 and 4, are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 dimerization, internalization of the EGF-receptor complex,
 induction of the tyrosine kinase activity, stimulation of cell DNA
 synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X00588; CA035240.1; -
 DR EMBL: U95088; AAB50863.1; -
 DR EMBL: U48722; AAC50802.1; -
 DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -

Query Match 45.0%; Score 3065; DB 1; Length 1210;

Best Local Similarity 49.2%; Pred. No. 1.6e-158;

Matches 623; Conservative 173; Mismatches 361; Indels 110; Gaps 25;

QY 11 LILALPFGAA--STVCTGTGTMKRLRPLASPTHTDMLRHLYGCGGVVQGLTLYPTN 68
 Db 14 LLAALCPASALEEKVKVCGGTGTSNKLTGCTGFEDHFLSLQRMFNNEVVLGLLEITYVQRN 73

QY 69 ASLEFLDIDQEVGVYUILLANOVROVPLQRLAIVRGCTQLFEDNYVALVDNGPLNNTTP 128
 Db 74 YDLSFKLTIOEVAQYVLIANTVERPLENIQIRGMYEENSVALAVSNVD----- 126
 QY 129 VTGASPGELRELQSRSTELIKGAVLIQRPOCYDPTLMKDIPIKNNQALATLIDTR 188
 Db 127 ---ANKTGKELPRNIOELHGAVERSNPNPALTGNVESTIOMRDISSDFLSNMSKPFQNH 183
 QY 189 SRACHPSPMCQSGRCWGESEEDCQSLTRVTCAGCA-CKGELPTDCCHEQCAACTGP 247
 Db 184 LGSQCKDPSCPNQSCWGAEEFNCKLTIKILCAQCSGRCGRKSPSDCCHNQCAACTGP 243
 QY 248 KHSPCLACHPNNSIGICELCPALVYNTNDTFESMNPGRRTYFGASCTYACPYNTLSD 307
 Db 244 RESDCLVRKFRDPAATCKDCCPLMLNPTTYQMVDNPFEGKYSFGATCVKCPERNVVD 303
 QY 308 VGSCTLVCPLHNOEVTAEQTCRCKSPCARVCYGLQMVIKANSKFIQTELE-FAG 366
 Db 304 HGSQVRACGADSYEM-EDGVKRCCKCEGRCRVKNGIDIGERK-DLSINATNIGHFKN 361
 QY 367 CKIFGSLAPSPSFDGDPASNTAPLQPEQLVFTLEITGYLYISAMPDLPDISVFQ 426
 Db 362 CTISGDLHLPLVAFRGDSFTHTPPLDPELDILKTKVEITGFLILQAMPENRTDLHAE 421
 QY 427 NLQVIRGRIILHNGAVSLTQGLGISMLGRSLRELSSGLALIHNNHTLCFHTVPMDQLF 486
 Db 422 NLEIRKRTQHQGFSLAVVSLNTLSGLRSLKEISDGVYISGNKNLCYANTINMKLF 481
 QY 487 RNPHQALLTANRDEDECVGEGSLACHQLCARHGCMGPGPTQCVNCSQFLRGCEVECRV 546
 Db 482 GTSGQCKTKIISNGENSCAKTGVCHALCSPGQWCEPRDVCNVRNREGVCDKCL 541
 QY 547 LOGIPREYVNAARCLCHEPCQONQSVYCFEPBADQCAAHYDPPCVARCSGVKP 606
 Db 542 LEBGPREFENSCFICHPCELPQAMNITCTGRPNCCIQAHYDIDPCHVCTCPAGWVG 601
 QY 607 DLSYPMKRPDEEGACOPPCINCTHSCVDLDDKGPARGASPLTSYSAVG---IL 663
 Db 602 ENNTL-VMKXADAGNHCHCHPCTCYGCGPGLEGPTNPRIP--STATGWGALLLL 658
 QY 664 VVVLGVVFGILIRROOKRKXTMRRLQETELVBLTPSGAMPQAOIRLKEQYIRAN 723
 Db 659 VVALGIG--LFMRRHIVKRTLRLLQERELVBLTPSGEAPQALRLIKETEFK-K 714
 QY 724 SKFQITEL-TVYKGIWIDGSEVVKIPVAILKRLRENTSPKAKETLDAVYVWAGSPYV 782
 Db 715 IKVLGSGAGTIVYKGLMIFEGEKVPIVAILKRLRENTSPKAKETLDAVYVWAGSPYV 774
 QY 783 SRLIGICLTSTVQVLTQMPYGCCLLDHYENRGRLSQDILNMCQIAKMSYLEDLV 842
 Db 775 CRLLGICLTSTVQVLTQMPYGCCLLDHYENRGRLSQDILNMCQIAKMSYLEDLV 834
 QY 843 HRLAARNTLVSPNHNKTTDGLARLIDIDTEHNAQGVKPYRMALLESILRRRTFHQ 902
 Db 835 HRLAARNTLVKTPQVKTIDGLAKLGAEEKEKHAAGKVPYRMALLESILRRRTFHQ 894
 QY 903 SDVMSYGVTVMEILTFGAKPYDIPAREIPDLLEKGERLPPOPITIVYIMYKCMWID 962
 Db 895 SDVMSYGVTVMEILTFGAKPYDIPAREIPDLLEKGERLPPOPITIVYIMYKCMWID 954
 QY 963 SECRPFRELVSEFSRMAADPQRFVVIQ-NEDLGPASPLDSTFYKSLLEDDMDMDLVDAE 1021
 Db 955 ADSRPFRELVSEFSRMAADPQRFVVIQ-NEDLGPASPLDSTFYKSLLEDDMDMDLVDAE 1014
 QY 1022 EYLVPOQGFPCDPAPAGAGVNHHRSSSTRSGGDLTLLEPSBEERAPSLPSEGA 1081
 Db 1015 EYLVPOQGF-----SSPSTSTPTLLSLISA 1040
 QY 1082 GSDVFDGIDGMAAGLQSLPHTDPSPLQRYSEDPTVLPSET--DGYAVLTPCSPOPEY 1139
 Db 1041 TSN--NSTVACIDRNGLSQSPFKEDSFQRYSSDPTALTDSIDDTFL-----PVPEY 1092


```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 (BY SIMILARITY)
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL)
FT CONFLICT 19 C -> S (IN REF. 2)
FT CONFLICT 539 539 C -> W (IN REF. 5)
FT CONFLICT 991 991 L -> F (IN REF. 4)
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6)
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2DF5 CRC64;

```

Query Match 44.6%; Score 3039; DB 1; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 4,2e-157;
 Matches 625; Conservative 167; Mismatches 368; Indels 114; Gaps 27;

```

QY 11 LLLALPFGAA--STVCTGTDMLRLPASPEHLDMRLHYOGQVQVQNTLEITYPTN 68
DB 14 LLLALCAAGALEBKVKCGGTNRLTQCTPEDHLSLQRYNNCEVGLNLEITYVQRN 73
QY 69 ASLSFLQDIQEVQVYLAAHNOVRVPLQRLRYVGTQLFEDNYALAVLDNGDEPLNTP 128
DB 74 YDLSFLKTIQEVQVYLALNTVERIPLNLIQIRGNALYENTYALALISN----- 124
QY 129 VTGASPGLEREQLSLREILKGVLIQRNPQLQVQDTILMKDI-----FHQNLALTLI 184
DB 125 -GTRRTGLRELPMKNIQELILGAVRSNNPILCMQDTIQMRDIVQVNFSSNMSMD- 180
QY 185 DTNRSRACHPCSPMKSGKSCWGESSEDCSLTRTVACAGCA-RCKGFLPTDCHQCGAAG 243
DB 181 -QSHPSCEPKCDPSCPNCSGSCWGEHCQKLTKIIICQCGSHRRKSGSPSCCHNQCAAG 229
QY 244 CTGPKHSCLLACLNHNHSGICEHNPALVTYNTPTFESMPNPEGRYFPGASCYTACPYNY 303
DB 240 CTGPRSDCLVCQKQDEATCKDTCPPMLNPTTYQMDVNPBEKYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTLVCPLEHNOEVTAEADGTORCEKSKPCARCVYGLGMQYIKANSKFEITEL 363
DB 300 VYTDHSCVRAAGPQYEV-EDGIRKCKKDDGCRKVCNGIGIETK-DTSLINATNIK 357
QY 364 -FAGCKTIFGSLAFPESEFDGPPASNTAPLOPEQLQVETLEITGYLYISAMPDLPDL 422
DB 358 HFXYCTAISGDLHLPLVAFKGDSEFTPTPDLPRELEILKTVKEITGFLIQAMPDMWTDL 417
QY 423 SVFQNLQVIRGRILHNGAVSLTLQGLGSMVLGASLELQSLALIHNNHLCFNTIVM 482
DB 418 HAFENLEIRGRITQHQOFLAVGLNITSLGSLSLKSIDGDIVLISGNNLQVANTIVM 477
QY 483 DQLEPNHQALLHTANPREDECEVGEGLACHQLCARHGWSPGPTQCVNCSQFLRQGEVE 542
DB 478 KKLFGTPQOKTKIMNNAEKCKKAVNHNVPILCSSEGCWGPBEPDVCSCQNSRGKECE 537
QY 543 ECRVLOGLPREYVVARHCLPCHPCOPONSVCYCFEALQCCVACAHYKDPPTCVARCS 602
DB 538 KGNILGEGFRFVNSSECIQHPCLPQAMNITCTTGPGNCCIQCAHYIDGPRCVATCPA 597
QY 603 GVKPDLSTMPIMKPEDEGACQPCPINCSTHSCVDLDKGPAPACQASPLTSIYSAVGL 662
DB 598 GIMGENNTL-VMKYADANNVCHLCHANCYGCAGPGLQGEVWSPSPKIPSTATIGVGL 656
QY 663 LVVVLGVVFGI-LIKRQOKIRKTYTRMLLQETELVEPLTPSSAMPNQOMRLKEQYIK 721
DB 657 LFTIV-VALGIGLWRRRHIVRKTRRLQERELVEPLTPSGEAPNOAHLILKETEFK 715
QY 722 ANSKFIGITEL-TVYKGIWIPDGENVAKPVAIKVLENTSPKANKELIDEAVYMGVGP 780

```

```

DB 716 -KIKVLGSAFGFVYKGLIPEGEVKIPVAIKELREATSPRANKELIDEAVYMAVSVP 774
QY 781 YSRLLIGLICLTSTVQLVQLMYGGLLHVENRGRCLSSQDLAMCQIANGMYLEVR 840
DB 775 HVCRLIGLICLTSTVQLVQLMYGGLLHVENRGRCLSSQDLAMCQIANGMYLEVR 834
QY 841 LVHRDLAANVLVYKPNHYKIDPGLABLDIDEREYADGKVPKMALESLIRPFT 900
DB 835 LVHRDLAANVLVYKPNHYKIDPGLABLDIDEREYADGKVPKMALESLIRPFT 894
QY 901 HOSDWSYGVTVLWELMTFGAKPYDGIPIAREIPDLLEKGERLPPEPCTIDYMIWVKCM 960
DB 895 HOSDWSYGVTVLWELMTFGSKPYDGIPIASDISILEKGERLPPEPCTIDYMIWVKCM 954
QY 961 IDSECRPRRELVSFESKAPRQCFVYIQ-NEDGPRASPLDSTYRLLEDMDGLVD 1015
DB 955 IDDSRPRRELILFESKAPRQCFVYIQ-NEDGPRASPLDSTYRLLEDMDGLVD 1014
QY 1020 AEEYLVPOQGFPCPDPAQAGAMVHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSE 1075
DB 1015 ADEYVILPOQGF-----NSPST-----SRTPILSL 1040
QY 1080 GAGSDVFDGDLGMAKGLQSLPTHPSEFLORYSBDPTVLPST--DGVAAPLTCSPOP 1137
DB 1041 SATSN-----NSTVACINRNGSCRYKEDAFLORYSDPTGAVTEDNIDDAFL-----FVP 1090
QY 1138 EYVNOPDVAPQPPSPREGPLPAAPRAGATLERATLSGKXGVVXKDVAFAGAVENPEYL 1197
DB 1091 EYVNO-SVPKPRAGSVQVFNHNPDLH-----AFGRDLHYQN--PSNAGNEPEYL 1139
QY 1198 -TPGGAAPQPPPPAPSPAFDNTYWDQ-----DP-----PERGAPSTFEKT 1240
DB 1140 NTAQ-----PTCLSSGFNSPALMTQKSHQMSLDNDPQYQDFFPKETKPRGIRKG- 1189
QY 1241 PTAENPEYLGIDVP 1254
DB 1190 PTAENAEYLRVAVP 1203

```

RESULT 6
 ERB4 HUMAN
 ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=9318974; PubMed=8383326;
 RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlson G.W.,
 Foy L., Neubauer M.G., Shoyab M.,
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
 RT epidermal growth factor receptor family".
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
 RA Klagsbrun M.,
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester".
 RL J. Biol. Chem. 272:26761-26768(1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BECAMELLULIN AND
 CC NTAR. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL, L07868; AAB59446.1; -
 DR HSSP; P11362; IFGK.
 DR Genew; HGNC:3432; ERBB4.
 DR MIM; 600543; -
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR Prodom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00263; EU_4.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.
 FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1168 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC).
 FT VARSLLIC 626 648 NGPSSHCIYPMQGHSTLQHA -> IGSSIEDCIGLMD
 (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE8095D88761 CRC64;
 Query Match 42.4%; Score 2887.5; DB 1; Length 1308;
 Best Local Similarity 44.5%; Pred. No. 7.1e-149;
 Matches 599; Conservative 184; Mismatches 389; Indels 175; Gaps 30;
 9 WGLLALLPPGAA-----STVCTGTMKLRIPASPEHMLMLRLHYGCCVVOGNNLELY 64
 8 WWSVLLVAAGTVPSPDSQSVCACTEKLSTSLDLEQYALKRYENCVCVMGNLEITS 67
 65 LPTNASLSFLDIOEVQGVYIAHQVROYVPLQRLVRGQVLPEDNYALAVLDNDGPLN 124
 68 IEHRDLSFLRSVEVGVYVALNORRYLPBLRLIRGKLEEDYALALFLNTRKDG 127
 125 NTPVYTGASPGGLRELQRLSTLEILKQGVLIQRNPOLCYDQTLIMKDIFKNNQLATLI 184
 128 NF-----GLOELGKMLTEILNGGVYVDQKFKCYADTIHQDQIVANPMSNLTIV 178
 185 DTNRSBACHPSPCKSGKSGESSEDCQSLTRVYVCGG-ARCKGLPTDCCHEQCAAG 243
 179 STNGSSGCGRCHKSCGTG-RCWGPTEHNCQTLTRVCAEQCDGRCGYGVYSDCCHEQCAAG 237
 244 CTGPRGSDCLACLFHNSHGICELCPALVTYNTDTEFSMPNREGRYFYGASCTYACPYNY 303
 238 CSGPRDTCFCACNPNFNSGACVTCGCPQTPYNNPTTFLDLENPNAKTYGAFCYCKRPHNF 297
 304 LSTDVGSSTLVCPHINQVYTAEDGTQCEKSCSPKACVYCYGLDMQYIKANSKRTGTELE 363
 298 V-VSSSSVCRAKCPSSKMEV-BENDIRKXKCDCTIDCPKACDGIQSGISMSAQTVSSNIDK 355
 364 FAGCKTIFGSLAFPIESFGDPPASNTAPLPEOLQVETLEITGVLYISAWDSIDPDS 423
 356 FINCKIKNGNLIPLVTHGHDPYNAIAPPEKLNFRYVREITELGINSQWPPNMTDS 415
 416 VFSNVLVIGGVVLSVGLSLILKQOGITISQFOSLKEISAGNIYIDNSNLCTYHTINMT 475
 424 VFQNLQVIRGILNNGAYSLTLQGLISWGLGSLRLPGLSGLALIHNNHTLCPVATVPM 483
 484 QLFNPHQALHTANRDEDECVEGALAQGLCARGHMGWGGPFCQVNCQFLRGSCVBE 543
 476 TLFSTINQRIYIRNRAENCTABGAVCNHLCSGDCGKGPPQCSGFRFSGRCTIS 535
 544 CRVIGQLPREYVNAHCLPCHPCOP-QNGSVTCFQPEADQCVACAHYDPPFCVAPCS 602

Db	536	CNLYDGEPRFENSGICVCECDPCCEKMEBGLLTHCHGPRGDNCTKSHKQDNCEVCPD	553
Qy	603	GKPEDLSMPFMKPFDEGACQCPRCINCHSCYDLDKCC-----PAQASPL	651
Db	556	GLQGANSF--LFKXVADPRCHCPCHNCCTQCGNCPRTSHDCLYPTWGTSTLPHAR--TPL	652
Qy	652	TSISAVV-GILAVVYGVVFGILLIKRROOKIKRYKMRPLLOETLVEPLTPSGAMNOA	710
Db	653	--IAAGVIGGFLIVITIGLTFAYAYVRKSIK--KKALRRL-ETLVEPLTPSGTAPNOA	708
Qy	711	QMRLLKEQYIKANSKFTIGTEL--TVYKGIMIDPGSNVKI PVAIKVLRNTSPKANKEILD	769
Db	709	QLRLIKETELK-RKYVLGSGAFGTGVKGIWVEGETVXIKPAIKLINETTGPKANVEFMD	767
Qy	770	EAYVWAGSGSYVSRILGICTSTVQLTQVLMPCYGLLDHRENRKRGSGSDLLNMCMQI	829
Db	768	BALIMASDHPLVRLGLVCLSTPTQLVQLMPHGGLLEVHEHNDTNGSOLLNMCVOI	827
Qy	830	AKGMSYLEDELVYHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGQVPIKMM	889
Db	828	AKGMVYLEERLVLVHRLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGQVPIKMM	887
Qy	890	ALLESLERRFTHQSDWSYGVTVWELMTGCAFPYOGIFAREIPLDKERLPPQPICTI	949
Db	888	ALBCHIRKFLTHQSDWSYGVTVWELMTGCAFPYOGIFAREIPLDKERLPPQPICTI	947
Qy	950	DVYIMVYKCMWIDSECRPFRELVSEFSRMADPORFVIONED-LGPASPLDSTFYRSI	1001
Db	948	DVYIMVYKCMWIDSDSPKFKELAAFSRMADPORVILVIGODRMLPSBPNDSKFFQNI	1007
Qy	1009	LEDDMGDLVDAEELYVPOQGFCCPDPAPAGAGMWHHRHSSTRSGGDLTGLEPSEE	1066
Db	1008	LDEEDLEDMDMAEELYVP--QAFVIRPP-----IYTSRARLDSNRS-----EIGHSPPA	1055
Qy	1069	EAPRS-----PLAP--SEGAGSDVDFDGLMGAAKGLQ	1093
Db	1056	YTPMSGQFYRDCGSGAAGQGSVPYRAPSTITPAPYAGATKTEIPDSCNGTLRKVY	1111
Qy	1100	SLPHHPDSPLOKRYSEDPYPLDS-----ETDGYVALTCSPPQPEVYNOPDYRPOPSP	1157
Db	1116	APHVQEDSSTORYSADPTVFAPEERSRGELDEEGYTMKDKPKQEYLVNE-----	1166
Qy	1153	REGPLPARPAGATLERATLTSPGNKYGVKVFAGCAVEMPEYLTITOGGAAPQPHPPA	1211
Db	1168	ENPFVSR-----KNGDLQ-----ALDPEYHNASNG-----PPYA	1196
Qy	1213	-----FSPAFDLYWYWDQDPPERGA--PST	1236
Db	1199	EDEVYNEPLYNTFANTLGKAEYLKKNILISMPERAKKAFQNDPVWNNHSLPRTSLQHPDY	1258
Qy	1237	FKGPT-----AENPEYL	1249
Db	1259	LOEXSTKYFYKONGRIPIVAENPEYL	1285
RESULT 7			
ERB4 RAT			
ID	ERR4 RAT	STANDARD:	PRT: 1308 AA.
AC	Q62956; Q922N7;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbs-4 precursor (BC 2.7.1.112).		
GN	ERB4 OR TYRO-2.		
OS	Rattus norvegicus (Rat)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
CX	NCBI_TaxID:10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Heart;		
FX	MEDLINE=9821155; PubMed=9553078;		
RA	Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,		

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP BIND 718 985 PROTEIN KINASE.
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 ATP (BY SIMILARITY).
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 P-T -> N (IN REF. 3).
FT CONFLICT 1080 1082 Y-T -> YR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;

Query Match 42.1%; Score 2868; DB 1; Length 1308;
Best Local Similarity 44.3%; Pred. No. 8,1e-146;
Matches 597; Conservative 191; Mismatches 398; Indels 162; Gaps 30;

1 MELA-ALGRWGLL--ALLPQASTQVCTGTDMKRLRLPASPETHLDMRLHYQGCGVQ 57
1 MKLATGLWVGSLVLAARTVQPSASQSCAGCTENKLSLSLEQYALAKRYENCEYVM 60
58 GNLELTYPTNASLFLQDIOEVGYVLIANQVQVPLQRLRVGTQLFEDNYALAVL 117
61 GNLEITSEHNHDSFLRSIREVGYVALNOFRYVPLENLRIRIIRGTQLYEPRYALAI 120
118 DNGDPLNNTTPTVGTASPGGLRELQLRSYTEILKGVLIQRPOLCYQDQTLIMQDIFKHN 177
121 LNYRKDNF-----GLOELGAKNTEILNGSVYVDQNKFLCYADITHWQIVRNPW 171
178 QLALLTIDITNRSRACHPCSPMKCSRCGSESEDCQSLTRTVCAAGC-ARCKPLPTDC 236
172 PSNMTLVSTIGSSGGRCCHKSTG-RCWGPTENHCQTLTRTVCAEQCDGRCGYPRVSDCC 230
237 HEQCAAGCTGPKHSDCLAFHNSGICELHCPALVYVNTDPESMNPGRRTTFGASCY 296
231 HRECAGSGSGEKDIDCFACMFPNDGACVTCQPTFYVNPPTTQLEHENAKEYTGAFCV 290
297 TACPYNYLSTDVSGCTIVCPILHNEVTAEDGTQRCCKSPCARVCGYGLMOYIKANSKF 356
291 KKCPHNFV-VDSSSCVTRACPSKMEV--EENGIRKCKRCTDLCPRACDGIOTSLMSAQTV 348
```

```

QY 357 IGITLSEFAGCKIRFGSLAPLRESHDGDPASNTAPLQPEOLVFETLEITNGLYISAMP 416
DB 349 DSSNDKFLNCKKINGNIFLVTGHSPPVNAIDALDEKLVNTRVREITGFINIQTPM 408
QY 417 DSLPLSVFQNLQVTRGRILHNGAVSLTLQIGISWGLSLRSRELSGLALIHNTHLCE 476
DB 409 PMMTDFSVSNLVTIGRVLVSGSLILILKQGGITSLQFQSLKEISAGNIYITDNSLVCY 468
QY 477 VHTVMDOLFRRPHALLHTANRREDECVGGGLACHQLCARGHWGSGPPQCNVCSQFLR 536
DB 469 YHTIMVTLTFSTVNRIVIRDNRAENCTAGWCMHLCSDNGCMGSGPDCLSCRSFSR 528
QY 537 GQECVEGRVLQGLREYVNAHCLCPHPEQCP-ONGSVTCFGEAQCVCVAHYKDPF 595
DB 529 GVICIESCNLYGGEREFENSGICECDQSGCKNEDGLTTHGGRPNCTKSHFKQGN 588
QY 596 CYARPSGVKPDLSYMPJWKFPEDEGACQPCPINCTHSCVLDLDCG-----PA 644
DB 589 CVEKPPDYLOGANSF--IFKYADODRECHPCHPNCTQCGNPTSHDCIYYPWTHSTLPQ 646
QY 645 EGRASPLTISAVV-GILLVAVVGVVFGILIKRQOKIRKYMRLLOETELVEPLTPS 703
DB 647 HAR-TPL--IAAGVIGLFLIVALLFAVTRKSKT-KKRALRFL-ETELVEPLTPS 701
QY 704 GAMPQAGMRLKEGYIKANSKFLGITEL-TVYKGIWIPGENYKIPVAIKVIRENTSPK 762
DB 702 GAPQAOQLRIKETELK-RVKVLGSGAFGVYGVIMVESETYKIVALKINLETTGPK 760
QY 763 ANKELIDRAYMAGVGSYYVRLIGTCTSTVQVLTQIMPYGCLLDHYRENRGLSGDL 822
DB 761 ANVEMDALIMASVDHVLVRLGVCLSPITQVLTQIMPHGCLLEVHEHKNIQSLL 820
QY 823 LNMQCIKAGMSYLEDVLRVARDLAARVLYKSPNHKITDFGLARLDLDEYVHADG 882
DB 821 LNMQCIKAGMNYLEERLVARDLAARVLYKSPNHKITDFGLARLDLDEYVHADG 880
QY 883 KVPITMALLESILRRRFTHQSDVSYGYTWELMTFSKAPYDGIIPAREIDLLEKGRLP 942
DB 881 KVPITMALLESILRRRFTHQSDVSYGYTWELMTFSKAPYDGIIPAREIDLLEKGRLP 940
QY 943 QPCTIDVYIMVCKMIDSECPREELVSESRMARDPQRFVYVQNEB-LQPASPLD 1001
DB 941 QPCTIDVYIMVCKMIDSECPREELVSESRMARDPQRFVYVQNEB-LQPASPLD 1000
QY 1002 STFYSLLEDDMDGLVDAEYLVPOQGFCCPDP-----APG 1038
DB 1001 SKFQNLDEEDLEDMDMDAEYLVQ-QAFNIPRTYSTRIDSNRSEIGHSPRPATPM 1059
QY 1039 AGMWHHHRSSSTRSGGDLTLGLESEBEARSPPLASEGASDVFDGLMGAAKL 1098
DB 1060 SSGQFVYDQGFATQGG--MPMYTATSTIPAPVA--QGAIAEMFDDSCCNGLRKP 1114
QY 1099 QSLPTHPSPLORESDEPTVPLPS-----ETDGYAPLTCSPQEPYVQVPRPQPS 1151
DB 1115 VVPHVQEDSSQYRSADPTVAPARNRRAALDEGGYMTPHMDRPEKQEYLNVE----- 1167
QY 1152 PREGRPLPARAGATLEBAKTLKSGKGVKVDVAFAGCAVENPEYLTPOGGAAPQHPRP 1211
DB 1168 --ENPFVSR-----KNGDLQ-----ALDNEEYSASAG-----PRK 1197
QY 1212 A-----FSPARDNLYYMDQDPERGA--PPS 1235
DB 1198 AEDRYVNEPLVLTFTNALGNAEYMKRSLLSVEBKAKAFDNDYMHSLPRSTLQHPD 1257
QY 1236 TFKGTPT-----AENPEYL 1249
DB 1258 YLOEYSTYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;
STANDARD; PRT; 1167 AA.
```

FT	DISULFID	237	245	BY SIMILARITY.
FT	DISULFID	241	253	BY SIMILARITY.
FT	DISULFID	256	265	BY SIMILARITY.
FT	DISULFID	269	296	BY SIMILARITY.
FT	DISULFID	300	311	BY SIMILARITY.
FT	DISULFID	315	330	BY SIMILARITY.
FT	DISULFID	333	337	BY SIMILARITY.
FT	DISULFID	504	513	BY SIMILARITY.
FT	DISULFID	508	521	BY SIMILARITY.
FT	DISULFID	524	533	BY SIMILARITY.
FT	DISULFID	536	553	BY SIMILARITY.
FT	DISULFID	566	569	BY SIMILARITY.
FT	DISULFID	560	577	BY SIMILARITY.
FT	DISULFID	593	615	BY SIMILARITY.
FT	DISULFID	618	626	BY SIMILARITY.
FT	DISULFID	622	634	BY SIMILARITY.
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	144	144	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	501	501	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	621	621	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1167 AA;	129934 MW;	4793E0749DC1D55A CRC64;

Query Match 38.4%; Score 2614.5; DB 1; Length 1167;
 Best Local Similarity 44.4%; Pred. No. 3.7e-134;
 Matches 564; Conservative 167; Mismatches 397; Indels 141; Gaps 31;

QY	4	AALCRWGILLALPERGAAT----	QVCTGTDMKRLRPASPTHIDMRHLYGSCOVQGN	59
DB	8	AALQD--LILVLISRCSTDPDRKTCVQSGNSNQTM--	LDNHLYLKKKKMYSGCNVLEN	62
QY	60	LELYLPTNASTLTDIOIQVQGVLLAHNQYQVRLRYANGTOLFEDNVALAVLDN	119	
DB	63	LEITYQENQDLSLQSIQVQGVLLIANNEVSTIPLVNIIRLIRGQNIYEGNFTLLYMSN	122	
QY	120	GDPLNNTPTVYGAAGGLREIQLSTLEILKGVLLIQRPOLQVDTILMKDIPHXNQQL	179	
DB	123	YQK-NPESP--DYVQVQLKQLQLSNLEILLISGVKYSNPLLCVETILNMWDYDKTSPN	179	
QY	180	ALTLIDNRSBACHPCSPMKGSRHCWSESSDCQSLTRTVCAAGC-ARCKGPLYTDCHE	238	
DB	180	TWNLIPIHAFEOCKCDHGCVNGSCMAPRGCHQCKFYKLLCAEQCNRRGRGPKRIDCCNE	239	
QY	239	QCAAGCGPKNKSDCLACLHFNHSGICGLHCPALVTYNTDFFESPNRNEGTYTGASCVTA	298	
DB	240	HGAGCGTGPATDCLACRDRNDQDTCQDTPPKYIDVISHQVANDNINIRYTGAAVCKE	299	
QY	299	CPVNYLSTDVQSGCTLVCPHNOETVTAEDGTQRCCKSKPAPRYCYGLGMQYIANSKFIG	358	
DB	300	CPSNVYVTE-GACVRSCSAGMLEVD-ENGRSKCPDQGVCPKYCDGIGISL-SNTIAVN	356	
QY	359	ITEL-ERAGCKKIFGSLAPFPEFDDGPASNTAPLQEOQLVFETLEITGYIYISAMPD	417	
DB	357	STNRSSTNCTKINGDILINRNSFEGDPHXIKGTMPBEHLMNLTYSKEITGYIVIMWPE	416	
QY	418	SLPDLVSFQNIQVIRGRILHNGAYS--LTQGLISIMLGARSLIELSGALIHNNHLCG	476	
DB	417	NMTLSLTFQNLLEIRGTTTSRGFSFVYVQVRLQMLGLRSLKEVSAQVILKNTQLCKRY	476	
QY	477	VHVPVMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMWGRPTQVNCSCQFLR	536	
DB	477	ANTINNRRLFRSEDSQISBYDART-----ENQCNNECSDEDCMGPRTMVCVSCLDHYR	529	
QY	537	GGSCVEECRLQLPREVYNARICLPCHEPCQPNQNSVTCFGEALQCYACAKYKRPFGC	596	
DB	530	GGCVASCNLLQGEPEPAQVDRCVQCHQECVLQYDTSILTCYGGPRANCSKSAHPQDQPC	589	


```
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDLCHMDITDMEDIYRDAEIVKNGR
SC -> GGFPMVSGLPPOADMDYLLDDPRLITLSASXK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLIT 560 560 E -> G (IN REF. 2).
FT CONFLIT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
```

Query Match 34.4%; Score 2341.5; DB 1; Length 1342;
Best Local Similarity 39.6%; Pred. No. 2.5e-119;
Matches 521; Conservative 197; Mismatches 460; Indels 137; Gaps 36;

```
OY 10 GLLALLPPGAA--STOYCTGTDMKRLPASPTHLDMLHLYOGQVVOGNILETYLPT 67
DB 11 GLFLSARGSEVNSGAVCPGTLINGLSTVGDAENQYQTLVKYVERCEVWGNLEIVLTH 70
OY 68 NASLFLQDIQENOGVYLIANNOVROVPLORLRYRGTOULFEDNYALALVDNDPPLNNT 127
DB 71 NADLSFLQMIREVTVGLVAVANNESTPLPLRLRYRGTOYDCKFAIFWV----LNYNT 125
OY 128 PVTGASPGGLRELOLRSTLTELKGVLIQRNPOLCYODTILMKDIFHKNNQLALTLDITN 187
DB 126 ---NSSHALRQLRLTQLTTELISGVYIEKNDKLCMDITDMEDIYRDR--AEIVVD 178
OY 188 RSRACHPCSPWCCKSRCKWGESSECCSLTRTVACGG--ARCKPRLPTDCCHECAGACTG 246
DB 179 NGRSCPCHEVCCK--RCKWGESSECCSLTRTVACGG--ARCKPRLPTDCCHECAGACTG 237
OY 247 PKHSDCLACLHFNHSGICEJHCPALVTYNTDTEFSMNPREGRTFGASCVTACPVNYLST 306
DB 238 PQDIDCFACRHFNHSGACVRCRQPLVYNKLTQLEBNPTKIQYGVAVASCPHFV-V 296
OY 307 DVGSGTIVCLPILHNOETVAEDGTORCEKSKPCARVCYGLQMOYIKANSKR--IGITEB- 363
DB 297 DQTSVRACPEDKMEVD--KNGLKWCCEPCGGLCPKACGEGT-----SGSRFQTVDSNIDG 350
OY 364 FAGCKTIFGSLAFPESEFDGPPASNTAPLOPELOVPELTLEETGLTYLISAMPDLSPLDS 423
DB 351 FVNCSTKLGNLDFITLNGDPWPKHKBALDPEKLNFRVTRREITGLNLTQSWPHEHNS 410
OY 424 VFQNLQVIRGRIHNGAVS--LTLQGLIGISWLGRLSRLDESSGLALIHNTHTLCPVYATPW 482
DB 411 VFSNLTITIGRSILYNRGFSLLIMKNLIVTSLGFRSLKEISAGRIYISANRQCYHHSLSNW 470
```

```
OY 483 DQLEFRNPOALLHTA--NRPEDECVGEGLACQGLCARGHWGPGPTQCVNCSQFLRGBCV 541
DB 471 TKVLGPTFEERLDITKRNPRDCAEBGVCPDLCSGAGCGPGQCLSKNYSRGVGV 530
OY 542 EECRVLOGLPREYVNAHRLCPHECOPONGSVTCFGEADQCVACAYKDPFCVARCPC 601
DB 531 THCNPLNGEPREFAHEACFSCHRECQMEGTATCNSGSDTCAQCHAFPDGPHCVSSCP 590
OY 602 SGVRKDLSTYMIKMFPEBEAGCQCPINCHTSCYDLDKCKPAEQR-----SPLTSVSA 657
DB 591 HGVLG--AKGPIYKPDVQNECRCHENCTOGCKGPELODCLGTLVLIGKTHLTMAVLT 648
OY 658 VVGLLVVVLGVVEGILIKRQOKR-KYTRRLLOETELVEPLTPSGAMPNOQWMLK 716
DB 649 IAG--LVVIFMGLGTFYMGGRITQKRAARLTLEGESETEPLDPS--EKANKYLARIK 705
OY 717 EGYTANSKFGITEB--TVYKGINIPGENVKIPVAIKVIRENTSPANKELIDEAYVMA 775
DB 706 ETELR-KLKVIGSGVFGVHKGVMIPEGESIKIPVCIKVIEDKSGRQSFOAVTDHMLAIG 764
OY 776 GVGSPYVSRILGICLTSTVOLVTQLMFYGLLDHYRENRGRIGSODILNMCQIAKMSY 835
DB 765 SLDNAHIVRLGLCPGSSLQVLVLYPLGSLDHRQALGQLLNMGVOIAKMY 824
OY 836 LEDVELVHRDLAARVLYKSPENHYKIDFGIARLLDIDETEHADGSKVPDKMALESIL 895
DB 825 LEEHGMVHRNLAAARVLLKSPQOVADFGAIDLPPDDQQLLYSEAKTITIKMALESIL 884
OY 896 RRRPETHQSDVMSYGVTVWELMTFAKXYDGIIPARETDLLEKGBRLFOPRTCTIDVYIM 955
DB 885 FGKTYHQSDVMSYGVTVWELMTFAEYVAGIURLAEVDDLKEGBERLAQPICTIDVYIM 944
OY 956 VKCMWIDSECRPRELIVSESRMARPPORFVYIOWNDLGRA---SPLDSTFYRLLEDD 1012
DB 945 VKCMWIDENIRPTKELANETRARPRPVIVYIKRS--GPGAIPGEPGILNKKLEEV 1003
OY 1013 DMGDLVDAEEVLVQOGFCPCDPAPAGVGVHHRHSSSTRSGGGLTLGLP--SEEAR 1071
DB 1004 ELEBELDLDDLLEAED-----NLATTTGSLALSLPVGTLNPRGS 1044
OY 1072 RSPAPSEGAGSDVFPDDDLGMAKGLQSLPTH--PSPLORYSEDPVPLP-----SET 1124
DB 1045 QSLSPSSGY-MPMNQNLBESCOESAVSSSERCPRVSLH-----PMRGCLASSESS 1097
OY 1125 DGYYA-----PLTCSPOPE-----YVNPQVRRPQPSPREGP----- 1156
DB 1098 EGHYTGSEAELOEKVSMCRSRSRSPRPGDGAHYSQSHSLTLPVLPISPLBEEDVN 1157
OY 1157 ---LPARPAGATLEAKTISP-GKXGV-----KDVFAFGAVENPEYLTQGGAAPO 1206
DB 1158 GYVMPDTHLKGTPSSRGLTSSVGLSSVLGTEBEDD-----EEYEYMNRRRHSP- 1208
OY 1207 PHPPAPSPAFDNLVYWD-----QDPPEGAPSTFGTPTAENPEYL 1249
DB 1209 PHPPRPSLDELGEYNDVGSDSLASLSTQSCPLPVPIMPITAGITTPBEDVBYM 1263
```

RESULT 10
ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
GN (c-erbB3).
DB ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	258	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
50	SEQUENCE	1339	147545	MM; 0AASF2402BPDFIE CXC64;

Query Match	33.3%	Score 2269.5;	DB 1;	Length 1339;
Best Local Similarity	40.0%	Pred. No. 2e-115;		
Matches	514;	Conservative 173;	Mismatches 438;	Indels 161;
Gaps	37			
Q	3	LAALCRWGLLALDPGAA---	STQVCTGTDIMKRLPASEPETHLDMLRHLVQCCVQVCGN	59
D	7	LDVLC---FLLSLARSGEMGNQAVCPGLTNGLSVGDADNQYQLTYLKYECCEVWKN		62
Q	60	LELTLYPTNASHSEFLDDIOEVQYVLLAHNQVQVLPORLIRYQTLFEDNYVALALDN		119
D	63	LEIVLTGHNADISFLQWIREVTGYVLVANKNEFSVLPPLNLRVVRGGVYDGKFAIFW--		120
Q	120	GRPLNNTPEVYGASGGIRELDQSLSTLEILKGGVLRQNPQQLCYOTILMKDIFHKMOL		179
D	121	---LWNT---NSSHLRQLKPLQLTEILSGGYLEKDKLCHMDTDIMRDLVRV---		170
Q	180	ALLTIDNRSRACHPCSPMCKSGSRWSESSDDQSILRTVCAGC--ARCKGPLPTDCEE		238
D	171	GAELIVKKNGANCPCEHEVCXG-RCMGPRDDCQILTKTICAPQCGRCFGPNPQCHD		229
Q	239	QCAAGCTGPKASDCLACIFHNHSGICELHRLVYNTDPEEMPRDEARYFGASCVTA		298
D	230	ECAGGCGRPOTDCCACRFYDSACVPCRPDELYVNKLTFQLEPRPHRTKIQGGVCAS		289
Q	299	CPYNTLSTDVQSCTLVCPCLHNOEYTAEDGTQRCSEKSKPCARVCYL--GMOYIRANSKF		356
D	290	CHNINV-VDQTFVYRACPRDMEVD-KHGKMCCEPGQZCPKACEGTSGSYQVTDASN		347
Q	357	IGITLEPAGCKKIFGSLAFPESEFDPSNAPLQPELOVETLEBIGYLYTISMP		416
D	348	ID----GVNCTKILGNMDELITSLANDPMHKIRALDPELVNFRVRRITGIYLIQSMR		403
Q	417	DSLPRLSFQNLQVIRGIRILANGYS-LTQIGIGIMLGRISIRELGGSLALIHNYTLC		475
D	404	PMNHMFVSFNLTTIGGSLNNGFSLIKMKNLVTSLSFRSKELISARVYISANQQLC		463
Q	476	FVHTVPMQDLFRNFHQAHLHTA--NRPEDECVSGELACHOLICARGHMGCPQCNCQSF		534
D	464	YHHSINMTNRLRGSEBERLDIKYDRPLGEGLAEGKVCDCPLCSGGCGMGPGQCLUSCNY		523
Q	535	LRGQCEVEGVLLGLPREYVNAHCLPCRPEDQPOGNSVTCFGRPADQVACAHYKRP		594
D	524	SREGVCTHCHFLDGGREEREFVHAQCSCFPECLLPEGISTGNGSGSDCARCAHFRGP		583
Q	595	FCVACPSGVKPRDLSYPIWKFPDEBGAQCPCEINTHSC--VDLDDKCAPAQRASPLT		652

```

DB      584 HCVNSCPHGLIG--AKGPITYKYPDAONECRPCHEHNTCGNGPELQJCLGQAEVLMSKPH 641
QY      653 SIVSAVGIILLVWVGAVFGLIKRROCKIR-KYTMRELLQETVELVPLPSSGAMPQAO 711
DB      642 LVIANTVG--LAVILMIGSSFLYWRGRILQNKAMRRYIERGESIEPLDPS-EKANKYL 698
QY      712 MRILKEQYIKANSKFEIGTEL-TYKGIWIPDGNVIPAIAKYLRENTSPPKNEILDE 770
DB      699 ARIFKETELR-KLKVLGSGVGFTVHKGIWIPEGSIKIPICINTEKSGRQGFQAVTGH 757
QY      771 AYWAGVSPVYSRLIGLITSTVQVLTQMPVCCLDHYRENRGRIGSDLLNMCQIA 830
DB      758 MLAVGSLDHAHIVRLGLCPSSSLQVLTQVLPISLLDHYKHQHEITGLPOLLNMCQVQIA 817
QY      831 KGMSTLEDVRLVHDLAARNVLRKSPNHVKTITPGLARLLDIBETEXHADGVKPKMA 890
DB      818 KGMYLEEHSVNHDLNARNVLRKSPNHVKTITPGLARLLDIBETEXHADGVKPKMA 877
QY      891 LESILRRFTHQSDVSVYGVTVWELMTFGAKPYDGIARSIPLLEKGERLPQPICTID 950
DB      878 LESIHFGKYTHQSDVSVYGVTVWELMTFGAEPAVGLRLAEIPDLEKGERLAQPOICTID 937
QY      951 VYMIWVKCMIIDSECPREFELVSEFSRMAPDQRFVIONEDLGPASPLDSTFYSLIE 1010
DB      938 VYMWVKCMIIDENIRPTFKELANEFTRMADPFRVIVIKRAS-GPQTP--PAAPESVLT 994
QY      1011 DDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVWHRHRSSTRSGGDLTLGLEPSEE-- 1068
DB      995 TKEL-----QEAELEPEL-----DLDLDLAEHEGL 1020
QY      1069 -----EAPRSFLASEG-----AGSDVFDGDLGMAKKGITQSLP 1102
DB      1021 ATSLGASLSLPTGLTPRRGQSLSLSPSSGMPMNQSLSEACLDASVLCGRQFSPRIS 1080
QY      1103 THDPSPLQRYSEDPVLPSETDGVY-----APL-----TC-----SPOPE-----YVNO 1142
DB      1081 LH-PIPRGR-----PASSESECHVIGSEAELOEKVSRSRSRSPRRPDSAYHQ 1132
QY      1143 PDVAPQPPSPREGP-----LPAPPAQATLERAKTISP-GKNGVY-----KDV 1184
DB      1133 RNSLITPTPLSPFGLEEDONGVYMDTLRGASASSREGTILSSVGLSGTEDEED- 1191
QY      1185 FAFGAVENPEYLLPOGGAAPQPP 1210
DB      1192 -----EVEYEMRKXRGSP-PRPP 1209

```

RESULT 11

```

EGFR_DROME
ID EGFR_DROME STANDARD PRF: 1426 AA.
AC P04413: O61601, Q9W2G0, P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DB Epidermal growth factor receptor precursor (EC 2.7.1.112) (EGFR)
DB (Gurken receptor) (Torpodo protein) (Drosophila relative of ERBB).
OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Mandibulata, Pancrustacea, Hexapoda;
OC Insecta, Pterygota, Neoptera, Endopterygota, Diptera; Brachycera;
OC Muscomorpha, Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=84350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RP [2]
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schjeter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkley;
RA MEDLINE=20196006; PubMed=10731122;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Abil J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Bolhaver S.,
RA Botkova D., Botchan M.R., Bouck J., Brockstein P., Brotier P.,
RA Butts J.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dup E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Huck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclev J.M.,
RA Palazzolo M., Peltan G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gblsh R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RP STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983222;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wenworth D.;
RA "A Drosophila genomic sequence with homology to human epidermal

```

RT growth factor receptor";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "there must be 50 ways to rule the signal: the case of the Drosophila
 RT EGF receptor";
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMIOSEBROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAA51462.1; -
 DR EMBL; K03417; AAA51460.1; -
 DR EMBL; K03416; AAA50965.1; -
 DR EMBL; K03418; AAA51461.1; -
 DR EMBL; AF109077; AAD26134.1; -
 DR EMBL; AF109078; AAD26132.1; -
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -
 DR EMBL; AE003454; AAF46732.1; -
 DR EMBL; X02293; CAA26157.1; -
 DR EMBL; X78920; CAA55523.1; -
 DR EMBL; X78918; CAA55521.1; -

DR EMBL; X78919; CAA55522.1; -
 DR PIR; A00640; GQFE.
 DR HSP; P11362; IEGK.
 DR Flybase; FBgn003731; Egf.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRODOM; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TIR; 1.
 DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 27.7%; Score 1887; DB 1; Length 1426;
 Best Local Similarity 32.0%; Pred. No. 1e-94;
 Matches 458; Conservative 187; Mismatches 434; Indels 352; Gaps 44;
 QY 24 QVCTGTDMKRLPASPETHLDMLHLYQCGVQGNELTYLFT-NASLFLDIDIEVQG 82
 DB 100 KICIGTSRSLSPSNKEHRYNRDRYNTCTYVDGNLKTWLPENMLDLSFLDNIRELVG 159
 QY 83 YVLAHNOVRVPLORLRIYRGTOLF-----EDNYALVADNDGDPINNTTPVGA PGGL 137
 DB 160 YILSHVDYKVVYPRKQITIRGRLTSLSYEEKYALFV-----TYSKM 203
 QY 138 RELQRLSLTELKGVLIQGNPOLCYQDITLMDIFPKNNQALTLIDNRSRACHPCSP 197
 DB 204 YTLRIPLRLDVLNQGVFHNHNNVLCHEMTIQSGEIVSNGTDAYYNDFTAPERECPKCHE 263
 QY 198 MCKSRMCGESSSEDCQSLTRTVCAAGCA-RCKGRPLPTDCHEQCAAGCTGPRHSCLAC 255
 DB 264 SCTHG-CWGGSPKNCQKFSKLTSPQACAGRCVGPAPRECHLFCAGGCTGPTQKDCIAC 322
 QY 256 LHFHSGICEHLCPALVTYNTDTFESMPNDEGRYTFGASCVTACPNYLTSDVGSCTLYC 315
 DB 323 KNPFDEAVSKSECPMKKNPTTYVLETNEGKAYAGATGVKCP-CHLLRDGACVRSQ 381
 QY 316 PLHQEVTYADGTCRCCKSPCARVYCGMGMYITRANSKFQITEL-----EPAGCKK 369
 DB 382 PQDKDKGGE-----CVPKNGPCPKTC-----PGVTYLAHGNIDSFNCTV 422
 QY 370 IFGSLAFLPESFDG--DPAQNTA-----PLOGEQVFEETLEITGYIYSAMPSLUD 421
 DB 423 IDGNIRILDQTFSPQVYVANYMGRIYPLDERREVSSTVEITGYINISGTHQFQN 482
 QY 422 LSVYQNLQVIRGLIHNGAY-STLLQGLISWGLRSLELGSGLALHNNHNLCEVHTV 480
 DB 483 LSYFRNLLETIHGQLMESMFALAIVKSLYSLENNLKOISGGSVAVIOHNNDLCVSN 542
 QY 481 PMDQTFRNPQALHTNRPDEDCVSGLAGLCLARGCWGPGTQCNCGQPLRGQCE 540
 DB 543 RWPATIQEPQKVMVNMENLRADLCENKGTICSDQCWEDGCGWAGTDQCLTKNFNNGTC 602
 QY 541 VEEGRVQLGLPREVYVNRHGLPCHPCEQPNQNSVTCGPEADQCVACAHYKDPFCVARG 600

```

DB      603 IADCGYISNAVK--PDNRFCIKCHEPCR-----TCNAGADGHCECHVAVDGGHCVSEC 654
QY      601 P-----SGVK-----PDL 608
DB      655 PAKKXNDRCVCECHATCDGCTGPTDGTIGIGACTTCCTAATINDATVCKCLDKDKCPD- 713
QY      609 SYMPTMKF--PDEGACAP-----CPI-----NCTH----- 632
DB      714 GY--FWEYVHPQEGSLKPLAGAVCRKCHPLCELCCTVYGHQVCSCTHYKREQCET 771
QY      633 -----SC-----VDLDKG----- 641
DB      772 ECPADHYTDEQRECFQRHPECNGCTGAGADCKSCRNFKLFDEANETGPGYNSTMFNCTS 831
QY      642 -CPAEOR-----ASPLTS-----IVSAVGLLVVYGVVFGI 673
DB      832 KCPLEMRHNYOYTAIGPYCAASPSSSKITANDVMNFIITGAVLVPTICICV--T 889
QY      674 LKRRQOKIRKYT--MRLLQTELEVPLTPSGAMPQAMRIKEQYIKANSKFIGITE 731
DB      890 YICROKAKKETVMTWALSGCEDSEFLRPSNIGANLCKLRIVKDAELRKGG-VLGMGA 948
QY      732 L-TYKKGWIPDGENVKIPVAIKVRENTSPKANKELDEAYVWAGSPYVSRLGICL 750
DB      949 FGRVTKGVAVPGEENVKIPVAIKELKSTGAESSEFLREAYVMASEHVNLKTLAVCM 1008
QY      791 TSTVOLVTLMPYGLLDHVENRGRILSGODLLWCMQIAKMSYLEDVLRHDLAARN 850
DB      1009 SSQMMILITQMLPGLGLDYVNNRDKIGSKALLMWSQIAKMSYLEKLVHDLAARN 1068
QY      851 VLKSPNHVKITDFCLARLLDIDETIYADGKVPKIMMALESILRRRFTHQSDVMSYGV 910
DB      1069 VLVQPSLVKLTDFCLAKLLSSDSNEYAAGKMPKIMLAECIRNVFTSKSVMAFVG 1128
QY      911 TWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICIDIVYIMVCMIDSECRPRR 970
DB      1129 TIMELLTFGQRPHENIPAKDIPDLIEVLKLEQELISGLDYICLLSCMHLDAMRPTFK 1188
QY      971 ELVSEFSRMARDPQRFVVIQNEIDLG--PASPLDSTFRSLLEDD--DMGDLVDAEYLV 1025
DB      1189 QLTTFEAFAPDPRGYLAIPGDKFTRLPA-----YTSQEKDLIRKLAPTDGSAIA 1241
QY      1026 PQGFFPCDPAPAGSGMWHHRSSSTRSGGDLTLGLEPSEBAP-----RSLAPSE 1079
DB      1242 KPDDYLOPKAAPGS-----KRTDCT-----DEMPKLNRYCKDPSNKS 1280
QY      1080 GAGSDVFDG---DLGMGAAGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQ 1136
DB      1281 STGDDERDSARVGVGNLR-----LDLPVDEDDYLMR--TCQEG 1318
QY      1137 PEYVNPQPVPRQPPSPRPGPLPAPRPACATLERAKTUSPGKNGVYKDVFAFGAVENPEY 1196
DB      1319 PNNNNNN-----NPNQNNMAAVGAAVYV-----DLIGVPVSVNPEY 1357
QY      1197 L-----TPQGAAPQPH-----PPPAFSP-AFDNLVYMD 1224
DB      1358 LINAQTLGVSESPITQITIGIPWNGCGTMEVKKPMGSEPTSSDHEYND 1408

```

```

RESULT 12
ERBB_ALV
ID_ERBB_ALV STANDARD; PRT; 634 AA.
AC_P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=8522822; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Cullen T.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.",
RL Cell 41:719-726 (1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
CC EMBL: M10066; AAA48763.1; ALT_INIT.
CC PIR: A00643; TVCHLV.
CC PIR: B00643; TVFVLV.
CC HSSP: P1362; IFGK.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; pkinase; 1
CC PRINTS: PR00109; TYRKINASE.
CC PRODOM: PD000001; Euk_pkinase; 1.
CC SMART: SM00219; TYRK; 1.
CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE: PS50011; PROTEIN KINASE DOM; 1.
CC TRANSFAS: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 24.5%; Score 1669.5; DB 1; Length 634;
Best Local Similarity 50.8%; Pred. No. 2,4e-83;
Matches 360; Conservative 79; Mismatches 144; Indels 125; Gaps 19;

```

```

QY 587 CAHYKDPFCVAPRSGVQKPLSMPLMKRPPDEGACGPPICTHSCVLDLCKGPABQ 646
DB 3 CAHPIDEPHCVKACPAVLGBNDTL-VKXADANAYCQDCHPNCCTGCGGEGCP--- 58
QY 647 RASPLTISVAVV-GILLVVLGVVPGILIKRQOKIRKYTRRLLOETELVPLTPSGA 705
DB 59 NGSKTPSIAGVVGGLCLVAVVGLIGLIVARR-HIVRKETLARLLQSEELVPLTPSGE 117
QY 706 MPNQAQRILKEQYIKANSKFIGITEL-TYKGIWIPDGENVKIPVAIKVRENTSPKAN 764
DB 118 APNQAHRILKETEFK-KVAVLSGAFGYKKGLMIPGKVKIPVAIKELREATSPKAN 176
QY 765 KEILDEAVVWAGVSPYVSRLGICLTSITVQLTQMPYGCGLDHYRENRGRILSGODLLN 824
DB 177 KEILDEAVVWASVDNPHVCGLLGICLTSITVQLTQMPYGCGLDHYRENRKNGISGYLLN 236
QY 825 WCMQIAKMSYLEDVLRHDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGGRV 884
DB 237 WCMQIAKMSYLEDVLRHDLAARNLVKSPNHVKITDFGLAKLLGADEKEYHAGGRV 296
QY 885 PIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPP 944

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds

(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 6815

Sequence: 1 MELALCRWGLLALPRGA.....TFKGTPTAENPEYLGDVPEY 1255

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phase:*
10: SP_plant:*
11: SP_prodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_rviro:*
16: SP_bacteriophage:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6129	89.9	1259	6 018735	018735 canis faml
2	3065	45.0	1209	1 090X70	090X70 rattus norv
3	3036	44.5	1210	11 09EP98	09EP98 mus musculu
4	2656	39.0	1165	13 09YH40	09YH40 xiphophorus
5	2615.5	38.4	1137	13 09W6F6	09W6F6 gallus gall
6	2225	32.6	1328	13 P79754	P79754 fugu rubrip
7	1971.5	28.9	1433	5 09B1H9	09B1H9 anopheles g
8	1871	27.5	419	4 09UK79	09UK79 homo sapien
9	1739	25.5	367	11 08R2X1	08R2X1 mus musculu
10	1697.5	24.9	412	4 08WV70	08WV70 homo sapien
11	1640	24.1	729	15 08E712	08E712 avian rous-
12	1638	24.0	567	15 08E714	08E714 avian rous-
13	1581.5	23.2	962	15 08E495	08E495 avian eryth
14	1573	23.1	545	15 08E468	08E468 avian eryth
15	1506.5	22.1	655	11 09WVFS	09WVFS mus musculu
16	1490.5	21.9	643	11 09ERV6	09ERV6 mus musculu

17	1211	17.8	1193	5 09Y1X8	09Y1X8 ephydaria f
18	1146.5	16.8	1368	5 023821	023821 caenorhabdi
19	1134	16.6	1717	5 026566	026566 schistosoma
20	1126	16.5	527	13 090836	090836 gallus gall
21	1001.5	14.7	478	11 09E8E0	09E8E0 rattus norv
22	942.5	13.8	599	13 09E8H2	09E8H2 gallus gall
23	906	13.3	165	4 014256	014256 homo sapien
24	887	13.0	176	11 0923V5	0923V5 rattus norv
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 08SZW1	08SZW1 drosophila
27	754.5	11.1	311	13 099162	099162 xiphophorus
28	734	10.8	331	13 09BUD7	09BUD7 homo sapien
29	723	10.6	149	6 09B6G6	09B6G6 oryctolagus
30	706.5	10.4	1362	12 09PVZ4	09PVZ4 xenopus lae
31	688	10.1	1671	5 09NJV5	09NJV5 biomphalaria
32	658.5	9.7	1368	13 08UW85	08UW85 paratichthy
33	650	9.5	1418	13 093457	093457 scophthalmu
34	646.5	9.5	1369	13 08UW86	08UW86 paratichthy
35	626	9.2	1358	13 073798	073798 xenopus lae
36	624.5	9.2	1472	5 09U5A8	09U5A8 bombyx mori
37	617	9.1	1412	13 08UW84	08UW84 paratichthy
38	607	8.9	1245	13 09YGH8	09YGH8 scophthalmu
39	605.5	8.9	1418	13 08UW83	08UW83 paratichthy
40	595.5	8.7	2144	5 09VD94	09VD94 drosophila
41	583	8.6	1371	11 09QVW4	09QVW4 rattus sp.
42	565	8.3	987	11 091YM0	091YM0 mus musculu
43	563.5	8.3	1091	4 09UMQ4	09UMQ4 homo sapien
44	562	8.2	935	4 096135	096135 homo sapien
45	562	8.2	987	11 099WR2	099WR2 mus musculu

ALIGNMENTS

RESULT 1
018735 PRELIMINARY; PRT; 1259 AA.
ID 018735
AC 018735
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erdb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_taxid=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erdb-2 from canine mammary gland."
DR EMBL; AB008451; BAB23127.1; -;
DR HSSP; P11362; IFGX.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9CA4CD46 CRC64;

Query Match 89.9%; Score 6129; DB 6; Length 1259;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 1132; Conservative 46; Mismatches 75; Indels 8; Gaps 4;

QY 1 MELALCRWGLLALLPGASTOYCTGTDMKLRPASPEHLDMRLHYGCCVVOGNT 60
 DB 1 MELAMCRWGLLALLPGAGAGTCTGTDMKLRPASPEHLDMRLHYGCCVVOGNT 60
 QY 61 ELTYLPTNASLFLDIOEVGVYLAHNOYQVLOGLRIVRGTOJFEDNYALAVLDNG 120
 DB 61 ELTYLPANASLFLDIOEVGVYLAHNOYQVLOGLRIVRGTOJFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLREJLRSLTEILKGGVLIQRNPOLCYODTILMKDIFKKNOLA 180
 DB 121 DPLEGGIPAPGAAGGLREJLRSLTEILKGGVLIQRSPOLCHOOTILMKDIFKKNOLA 180
 QY 181 LTLIDNRSRACHPGSPMCKGSRCKGSESEDCOSITRTVACGACRCGRLPTCCHEOC 240
 DB 181 LTLIDNRSRACHPGSPMCKGSRCKGSESEDCOSITRTVACGACRCGRLPTCCHEOC 240
 QY 241 AAGCTGPKHSDCLAHFNHSGICEHCPALVTYNTDFFESMPNDEGRYTFGASCTACP 300
 DB 241 AAGCTGPKHSDCLAHFNHSGICEHCPALVTYNTDFFESMPNDEGRYTFGASCTACP 300
 QY 301 YNLTSTVSGCTLVCPRLHQETADGQRCCKGSKPARVCYGLMOYIKANSKFTGIT 360
 DB 301 YNLTSTVSGCTLVCPRLNQETADGQRCCKGSKPARVCYGLMOYIKANSKFTGIT 360
 QY 361 ELEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPQOLVFPTLEETGYLYISAMPDLP 420
 DB 361 IOEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPQOLVFPTLEETGYLYISAMPDLP 420
 QY 421 DLSVFONLOVINGRILHNGAYSLTQGLGISWLGIRSRIRLGGSLALIHNTLCPVHTV 480
 DB 421 NLSVFONLRVINGRILHNGAYSLTQGLGISWLGIRSRIRLGGSLALIHNTLCPVHTV 480
 QY 481 PMDOLFRNPHOALLHTANRPEDECEGEGACHOQCARGHMGPRPTQCVNOSOTLRQEC 540
 DB 481 PMDOLFRNPHOALLHTANRPEDECEGEGACHOQCARGHMGPRPTQCVNOSOTLRQEC 540
 QY 541 VEECRVLOGLPREYVNAHRLCPCHPECPONGSVTCFGEADQCAVAHYKDPFCVAC 600
 DB 540 VEECRVLOGLPREYVNAHRLCPCHPECPONGSVTCFGEADQCAVAHYKDPFCVAC 600
 QY 601 PSGVCPDLISYPMIKFPEDEGACQPCPINCCHSCVDLDDKCCPAEQASPLTISVSAVNG 660
 DB 601 PSGVCPDLISYPMIKFPEDEGACQPCPINCCHSCVDLDDKCCPAEQASPLTISVSAVNG 660
 QY 661 ILLVAVLVGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAOMRLKXQYI 720
 DB 660 ILLVAVLVGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAOMRLKXQYI 720
 QY 721 KANSKFIGITEL-YYKGIWIPDGSNNKIPVAILKVLNNTSPKANKELIDBAVMAGVGS 779
 DB 720 R-KVYLVGSGAGFYKGIWIPDGSNNKIPVAILKVLNNTSPKANKELIDBAVMAGVGS 778
 QY 779 PYVSRLLGICLTSTVQVLTQMLPYCCLDHRHRRGRGSGDDLNNKCVQJAKGMSYEDV 838
 DB 778 PYVSRLLGICLTSTVQVLTQMLPYCCLDHRHRRGRGSGDDLNNKCVQJAKGMSYEDV 838
 QY 840 RLVRDRLAARNVLVSPVHVKITDGLARLLDIDETEXHADGKVPIMKMALESILRRRP 899
 DB 839 RLVRDRLAARNVLVSPVHVKITDGLARLLDIDETEXHADGKVPIMKMALESILRRRP 898
 QY 900 THOSDVMASYGVTVMELMTFGAKPDGIPAREIPDLLEGEELPPRPCTIDVYIMWKCM 959
 DB 899 THOSDVMASYGVTVMELMTFGAKPDGIPAREIPDLLEGEELPPRPCTIDVYIMWKCM 958
 QY 960 MIDSECRPRELVESEFSRMAPDPORFVIONEDLGASPLDSTFYSLDEDDMGDLVD 1019
 DB 959 MIDSECRPRELVESEFSRMAPDPORFVIONEDLGASPLDSTFYSLDEDDMGDLVD 1018
 QY 1020 ABEYLVPQGGFFCQDPAPAGAGMWHHRSSSTSGGGDLTLGLEPSEBAPRSPAPSE 1079

DB 1019 ABEYLVPQGGFFCQDPAPAGAGMWHHRSSSTSGGGDLTLGLEPSEBAPRSPAPSE 1078
 QY 1080 GAGSDVPDGLDMGAAGKLGSLPHDPSFLQRYSEDDPTVPLPSETDGYVADLTGSPQPEY 1139
 DB 1079 GAGSDVPDGLDMGAAGKLGSLPHDPSFLQRYSEDDPTVPLPSETDGYVADLTGSPQPEY 1138
 QY 1140 VNQPPVRPQPSPREGRLPARPPGATLER-----AKTISPKXGVYKVDVAFGAVENP 1194
 DB 1139 VNQPPVRPQPSPREGRLPARPPGATLER-----AKTISPKXGVYKVDVAFGAVENP 1193
 QY 1195 EYLTPQGAAPQHPHPAPFSPAFDNLVYWDOPPERGAPSTFGTPTAENPEYLGLDVP 1254
 DB 1199 EYLTPQGAAPQHPHPAPFSPAFDNLVYWDOPPERGAPSTFGTPTAENPEYLGLDVP 1253
 QY 1255 V 1255
 DB 1259 V 1259

RESULT 2

Q9QX70 PRELIMINARY; PRT; 1209 AA.

AC Q9QX70; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Epidermal growth factor receptor.

GN EGFR.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_Taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FISHER; TISSUE=LIVER;

RX MEDLINE=90258688; PubMed=2342466;

RA Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;

RT "A truncated, secreted form of the epidermal growth factor receptor is

RT encoded by an alternatively spliced transcript in normal rat tissue.";

RL Mol. Cell. Biol. 10:12973-2982(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=FISHER; TISSUE=LIVER;

RA Petch L.A.;

RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=FISHER; TISSUE=LIVER;

RA Guttridge K.; Dawson T.L.; Earp H.S.;

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; M37394; A014008.1; -

DR HSSP; P1362; IFG.

DR Interpro; IPR000494; EGFR_L domain.

DR Interpro; IPR000719; Euk_Dkinase.

DR Interpro; IPR002174; Eurin-like.

DR Interpro; IPR001245; Tyr_Dkinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Pkinase; 1.

DR Pfam; PF01030; Recep_L domain; 2.

DR PRINTS; PR00109; TYRKINASE.

DR Prodom; PD000001; Euk_Dkinase; 1.

DR SMART; SMO0261; FU; 3.

DR SMART; SMO0219; Tyrc; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS0011; PROTEIN KINASE DOW; 1.

DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.

KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.

SEQUENCE 1209 AA; 134891 MW; 96587F6CC18773 CMC64;

Query Match 45.0%; Score 3065; DB 11; Length 1209;
 Best Local Similarity 49.6%; Pred. No. 4,5e-224;

Matches		634; Conservative	164; Mismatches	363; Indels	118; Gaps	30;
Qy	3	LAALCRMGILLALPPGA-ASTQVCTGDMKRLPASPETHLDMRLHYOCQVQGNLE 61				
Db	15	LAALCAAG-----GALBEKKVCOGTSNRNLTLGTFEDHFLSLQMFNNCEVLGNLE 66				
Qy	62	LTYLPFNASLFLQDIDQVGVLIANQVQVPLQGLRIYVGTQLFEDNYALAVIDNG 121				
Db	67	ITYVQRNYDLSEFLKTIQVAGVYLIANTYERIPLENIQIIRGALYENYALAVLSN-- 124				
Qy	122	PLNNTPTVGASPGGLRELOLRSLTEIKKGVLLQRPOLCYODTILMKDIPHKNOQL 181				
Db	125	-----YGNKTKGLRELPMRNQELIIGAVRSNNPILCNMETIMKRLIV-QDVFLSN 175				
Qy	182	TLIDTNRS-RACHPCSPCKGRCWGBESSDQSLRTVAGGA-RCKGPLPTDCHEQ 239				
Db	176	MSMDVQRHLTGCCPKCDPSCPNCSWGGRBENCQTLTKICAQCCSRRCRCRSPSDCHNQ 225				
Qy	240	CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMNPGRRTFGASCYTAC 299				
Db	236	CAAGCTGPRESDCLVCHRFREDATCKOTCPRLMLYNFTTQMDVNPBGKYSFGATCVKGC 295				
Qy	300	PYNVLSTDVGSCTLVCPLANQVTAEDGTQRCCKSKPCARVCYGLMQYIKANSKFIGI 359				
Db	296	PRNYVYTDHGS CVACGPDYEV-EEDGVSKCKCDGPRCKVCGIGIGBPK-DTISINA 353				
Qy	360	TELE-FACCKKIFGSLAPFESFDGDPASNTAPLOPQLOVFETLETITGYLYTSAMPDS 418				
Db	354	TINKHFKYCTAISDDLHLPLPAFGDSFTPTPDPRELEILITVKEITGFLIQAMPEN 413				
Qy	419	LPDLVSFQNLQVIRGRILHNGAYSLTQGLISWLGSRSLREKSGSLALHNHTHLCFVH 478				
Db	414	WTDLHAFENLEIIRGRKTHQGFSLAVVGLNTSLGSRSLKEISDDVITISGNRNLCYAN 473				
Qy	479	TVPMQDLFRNPHQALLHTANPEDECVGEGIAHQLCARGHCWGPPTQCVNCSQPLRQ 538				
Db	474	TINKKTLFGTQNKTKIMNNAEKDCATNHCVCPLCSSBGCGPPTDVCQNVSRGR 533				
Qy	539	ECVEECRYLQGLPREYVNAHRCULCPHECQPNQSVTCFEPDQCVACAHYDPFCVIA 598				
Db	534	ECVDKCNILBEPREFBNSEFCICCHECLPQTNKITCTGRGPNCKCKAHYVDGHCWK 593				
Qy	599	RCPGKVRDLSYPMKFPDEGAQOPCPINCTHSCVDLDDKDCPAEORASP-LTSIVA 657				
Db	594	TSPSGIMGENTTL-VMKFADANNVCHLCHANCTYGGAGPGLKNC-QQBGPRKIPSIAG 650				
Qy	658	VSGILLVVLGVVFGI-LIKRQCKIKTKYMRRLQETLVEPLTSGAMPNQOMRIK 716				
Db	651	IVGGLFTIV-VALGICLFMRKRLQVARKTLRLQLERELVEPLTSGEAPNQAMRLIK 709				
Qy	717	EQYIKANSKEFIGTEL-TYKGIWIPDGENVKIPVAIKVLRENTSPKANKETIDEAYVA 775				
Db	710	ETEFK-KIKVLGSAFGTVKGLMIPGEKVKIPVAIKELREKTSKANKELIDEAYVA 768				
Qy	776	GVSGPYVSRLLGLTSTVQLVQLMPYGGCLDHYENRGLCSQDLLNWCQIAKMSY 835				
Db	769	SVDPHFVCRLLGLCTSTVQLTQLMPYGGCLDYVREHKNIISQYLNMWCQIAKMSY 828				
Qy	836	LEDVRLVHRDLAARNVYKSPNNVKIKIDFGARLLIDENYVADGKVPIMWALESTL 895				
Db	829	LEDRVLVHRDLAARNVYKTPQVKTITDEALYLGAEEKETHAAGKVPIMWALESTL 888				
Qy	896	RRRTHQSDVMSYGVYVWELMTGAKPYDIPAREIPDLLEKGERLPQPICTIDVYIM 955				
Db	889	HRIYTHQSDVMSYGVYVWELMTGSKPYDIPASEISILEKGERLPQPICTIDVYIM 948				
Qy	956	VKCMWIDSECRPRELIVSEFSYARDPQRFVYIQ-NEGLGASPLDSFTFYSLLEDDM 1014				
Db	949	VKCMWIDADSRPKFRELIIEFSKARDPQKIVLYIQDDEMHLPSPPTDSFYALMEEDM 1008				
Qy	1015	GDLVDAEEYVLPQGGFFCPDPABAGAGMWHRRSSSTRSGGDLTLGLPESEEPAPSP 1074				
Db	1009	EDVVDADDEVILPQGGFF-----NSPST-----SRTF 1034				

Qy	1075	LAPSEAGSDVFDGDLGKAAGLQSLPTHDPSPLOHYSDEPTPLPSET--DSYVAPLT 1132				
Db	1035	LILSLISANSN-----STVACINRNGSCVWEDATLQYSSDPTSVLTEDNIDTFL---- 1086				
Qy	1133	CSPOPEYVQPDVAPQPPSPREGFLPAPRAGATLEBAKTLSPKNGVGVXDFAFGAVE 1192				
Db	1087	--PVPEYINQ-SVPRKPAQSVQNVVYHNPFLHP-----APGRDLHYQN--PHSAVS 1133				
Qy	1193	NPEYL-TPQGAAAPQHPHPPAFSPFNLYVMQO-----DP-----PERGAPPS 1235				
Db	1134	NPEYLNINQO-----PCLSSGFSALMIQKSHQMSLDNPDYQODFFPEAKPNG 1184				
Qy	1236	TFKGTPTAENPEYGLDVP 1254				
Db	1185	IFKG-PTAENAEYLRVAPF 1202				

RESULT 3		Q9EP98	PRELIMINARY;	PRT;	1210 AA.
AC	Q9EP98;				
DT	01-MAR-2001 (Tremblrel. 16, Created)				
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)				
DE	Epidermal growth factor receptor isoform 1.				
GN	EGFR.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxId=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;				
RA	Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,				
RA	Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,				
RA	Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,				
RA	Malhe N.J.;				
RT	"Comparative genomic sequence analysis and isolation of human and				
RT	mouse alternative Egr transcripts encoding truncated receptor				
RT	isoforms."				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J;				
RA	Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,				
RA	Schehl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,				
RA	Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,				
RA	Malhe N.J.;				
RT	"Comparative genomic sequence analysis and isolation of human and				
RT	mouse alternative Egr transcripts encoding truncated receptor				
RT	isoforms."				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
EMBL	AF275366; AAG28045.1; -				
EMBL	AF275364; AAG28045.1; JOINED.				
EMBL	AF275365; AAG28045.1; JOINED.				
EMBL	AF275367; AAG24386.1; -				
HSSP	P11362; IRGX.				
MGD	MG1:95294; Egfr.				
InterPro	IPR000345; CytC_heme_bind.				
InterPro	IPR000494; EGFR_L_domain.				
InterPro	IPR000719; Euk_pkinase.				
InterPro	IPR002174; Furin-like.				
InterPro	IPR002290; Ser_Thr_pkinase.				
InterPro	IPR001245; Tyr_pkinase.				
Pfam	PF00757; Furin-like; 1.				
Pfam	PF00069; pkinase; 1.				
Pfam	PF01030; Recep_L_domain; 2.				
PRINTS	PR00109; TYRKINASE.				
ProDom	PD000001; Euk_pkinase; 1.				
SMART	SM00261; FU_5.				
SMART	SM00220; S_TKc; 1.				
SMART	SM00219; TYKc; 1.				

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Receptor; Transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.5%; Score 1036; DB 11; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 7,3e-222;
 Matches 625; Conservative 166; Mismatches 369; Indels 114; Gaps 27;

11 LLLALLPGAA--STQCTGIDMKLRPASBETHLMDLRHLYOGQOVVQGLLETTYPTN 68
 14 LTLALCAAGALEEKVKVCGGTSNRLTQGTEDHFLSLQRMVNNCEVVLGNLETTYQRN 73
 69 ASLSFLQDIQEVQVLLAHNQVQVPLQRLIRVGTQLFEDNALAVLDNGDPLNNTTP 128
 74 YDLSFLKTIQEVAGVLLALNVTERTPLENLQILRGNALYENTYALALSN----- 124
 129 VTGASFGGLRELQLRSLTEILKGVLLIQNPQLCYQDTILMKDI----FKNNQLATLLI 184
 125 -YGTNRTGLRELPMENLQELILGAVFSNNPILCMQDTIQMRDLYQNVFMSNMSMDL--- 180
 185 DTRNSBACHPCSPMKGSKGSESEDCQSLTRYVACGCA-RCKGPLPTDCCHGQCAAG 243
 181 -QSHPSCKPCDPCPNCSGCGGGENCQKLTKIICAQCCSHRCGRSPSDCHNQCAAG 239
 244 CTGPKHSDCLACLHNHSGICEIHCPLATVNTDTPFESMPREGRTFGASCYACPYNY 303
 240 CTGPRESDCLVQCKQDEATCKDTCPRLMLYNPTTYQMDVNEBKISFGATCYAKCERNY 299
 304 LSTDVGSCTLVCPRLHNOEYVADGTQRCSEKSKPCARCYVGLGMQYIKANSKPIGITELE 363
 300 VVTDGSCVACGAPPYEV-EEDGIRKCKCKDGCRCXVKNIGIGERF-DTISINATNIK 357
 364 -FAGCKIIFGSLAFIPSEFGDPASNTAPLQPEQVETLEETGLVLYISAMPDSDL 422
 358 HFKYCTAISGDLHILVAFKGSFTRTPDLRELEIKYKELTGFLLLQAMPDMNTDL 417
 423 SVFONLQVIRGRILHNGAYSLTLOGISLWGLRSLRELGSGLALIHNNHLCFVHTVP 482
 418 HAFENLEIIRGTQHQGFSLAVVGLNITSLGRLSKEISGDDVYISGNNNLCANTINM 477
 483 DOLFNNHQALLHTANRPEDECEVGEGLACHOLCARHGWGPGPQCNGSQFRLRQCEVE 542
 478 KKLFGTPNCKTKIMNNRAEKCKCAVNVNCLCSGEGWGPEDPCVSCNVSRGSECEVE 537
 543 ECRVLQGLPREYVNAHCLPCHPEGOPONGSVTCFGEADCCVCAHYKDPFCVACRPS 602
 538 KNILLEGEPREVESECIQHPBELPQAMNITCTGAGPNCICAHYIDGPHCVKTCFA 597
 603 GYKPLSLTPIWKPFDEEGACQPCPINCTHSCVDLDDKCPAEQRPASPLTISAVVGI 662
 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAPRGQCGEVPWSPGKIPISATGIVGL 656
 663 LVVLGVVFGI-LIKRQOKIRKTYMRLLOETELVEPLTPSGAMPQAOAMRLIXQYIK 721
 657 LEIVV-VALGIGLPMRRHIVRKATIRLLODERLVEPLTPSGAPQAPALRLIKETERF 715
 722 ANSKFIGITEL-TYKGIWIPDGNVVKI PVAIKVARENTSPKANKELIDRAYVMAGVSP 780
 716 -KIKVLGSGAGFATYKGLMIEGEEKVKIPVAKELREATSPKANKELIDRAYVMASVDNP 774
 781 YVSRLLGI-CLTSTVQVLTQMPYGLCDHVENNGRLGSGODLLNMCMQILKNGSYLEDVR 840
 775 HVCRLLGICLSTVQVLTQMPYGLCDHVENNGRLGSGODLLNMCMQILKNGSYLEDVR 834
 841 LVHRDLAARNVLYVSPNVKLTDFGLARLLDIDETEHADGQKVPKIMWALLESILRRFT 900
 835 LVHRDLAARNVLYVTPGVKITTDFGLAKLIGAEKEVHAAGKVPKIMWALLESILRRIT 894
 901 HOSDVMYSGVTVMELMTFGAKFYGTIGARELPDLLEGELPOPPCTTIDVYIMYKCM 960

DB 895 HOSDVMYSGVTVMELMTFGSKPYDGIASDISLLEKGEPLPPQPICTIDVYIMYKCM 954
 QY 961 IDSECRPFRELIVSEFSMARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGLVD 1019
 DB 955 IDADSRPKFRELILEFSMARDPQRFVVIQODERMHLPSTFDSFYALNDEDEMDGLVD 1014
 QY 1020 AEEYLVPOQGFPCDPAPAGAGMWHHRSSSTSGGDDTLGLEPSEEPAPSPAPSE 1079
 DB 1015 ADEYLTPOQGF-----NSPT-----SRPILLSL 1040
 QY 1080 GAGSDVPFGDGLGMAKAGLQSLPHNDESPLORYSEDPVLPSET--DGVAFLPCSPQ 1137
 DB 1041 SATSN-----NSTVACINRNGSCRYKEDAFLORYSSDPTGAVTEIDNIDATL-----PVP 1090
 QY 1138 EYVNPDPVPPQPSFEGPLPAAPAGATTLERAKTSPGKNGVYKVFAGAVENDEYL 1197
 DB 1091 EYVNO-SYKRRPAGSVQNPVYHNPPLHP-----APGDLHYQN--PSNAGNBEYL 1139
 QY 1198 -TPQGAAPQHPHPAPFAPFADNLVYWDQ-----DP-----PERGAPSTFGKT 1240
 DB 1140 NTAQ-----PCLSSGFSNPAIYQKSHQMSLDNPDYQDFFPKETKPNGLFKG- 1189
 QY 1241 PTAENPEYLGLDVP 1254
 DB 1190 PTAENAEYLRVAPP 1203

RESULT 4
 Q9YH40
 ID Q9YH40 PRELIMINARY; PRT; 1165 AA.
 AC Q9YH40;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorphi;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxId=8086;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RA Altschamed J., Scharlt M.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.";
 RL Oncogene 16:1681-1690(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharlt M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53471; AAD10500.2; -.
 DR HSP; P11362; IFGK.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Dkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR001245; Tyr_Dkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PS00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Dkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_pos_ANCHORING; UNKNOWN_1.

DR PROSITE, PS00107; PROTEIN KINASE, ATP, 1.
 DR PROSITE, PS50011; PROTEIN KINASE, DOM, 2.
 DR PROSITE, PS00109; PROTEIN KINASE, TYR, 1.
 KW ATP-binding; Kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7f7EE38D8771A4E CRC64;

Query Match 39.0%; Score 2656; DB 13; Length 1165;
 Best Local Similarity 44.9%; Pred. No. 5,6e-193;
 Matches 574; Conservative 163; Mismatches 390; Indels 150; Gaps 33;

```

QY 1 METALCRWGILLALPPG-AAST-----QVCTGDMKRLPASPETHLDMRLHVGQGV 55
DB 4 LELLE-----LILLILISGRCCSTDPRAKVCQGSNQM--LDNHYLKMKKXSGCNV 56
QY 56 VQGNLELYLPTNASLFLQDIQEVGVYLLAHNQVQVPLORLRIYRGTLFEDNYALA 115
DB 57 VLENLEITYQENQSLFLQSIQEVGVYLLAHNEVSTIPLVNLRLRGNIIEGNFTLL 116
QY 116 VLNDGDLPLNTTPVYGAPGRLRELQSLTEILKSGVLLQRPQLCYQDTIMKDI FHK 175
DB 117 VMSNYQK-NPESP--DYVQVGLKQQLSLNLEILSGGVKSHNPLLCNVEITIMMDIVDK 173
QY 176 NNOALTLIDITNRSPACHPCSPMCKGSRGWESSESDQSLTRTVCAQGC-AQCKPLPTD 234
DB 174 TSNPTNMLIPHAFERQCKDPCGVNCSMAFGCHQKTKLIDCAEQCNRRCKGRXPID 233
QY 235 CCHQCCAAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREGRYTFGAS 294
DB 234 CCHQCCAGGCTGPRATOCACRDFNDGCTKCTPPKIVDIYSHQVDPNFIKXFFGA 293
QY 295 CVTACPVYVLTQVSGCTVPCPLHNOEVTAEEDGOREKSKCAVVCGLM-----QYI 350
DB 294 CVKCEPSNYVTE-GACVRSASAGMLEVD-ENGKRSCKPCDGVCPKVCQIGIGLSLNTI 351
QY 351 KANSKFTIGITELEFAGCKKIFGSLAFLPESFDGPPASNTAPLOPEOLQVETLEITGYL 410
DB 352 AVNSTNIG---SFSNCKTINGDIILNRNFEGBPHYKIGPMDPEHMLLTYKELTGL 407
QY 411 YISAPDSLPLDLSFQNLQVIRGRLIHNGAYS-LTIQGLISLIGLSRISREIAGSLALIH 469
DB 408 VIMKPMNMTSLSFQNLLEIRIGTTFSGFSFVVQVSHQMLGSSLKEVASAGNVILK 467
QY 470 HNTHLCEVHTVPMQDLFRNPHQALLHTANRPEDECVGEGELAQHLCARHCWCGPPTQCV 529
DB 468 NTPQLRYASTINMRRLPRSEDSQISEVDART-----ENQTCNNESESGCWGPGTKVCV 520
QY 530 NCSQFLGQCEVEECRVLQGLPREVYNAHCLPCHEPCOPQNGSVTCFQPREADQCAQH 589
DB 521 SCHVDHGRGVASCNLLQGEPREADVGRCVQCHOECVLQGTDSLTCYGPFGANCSKCH 580
QY 590 YKDPFCVACRCPSPGVKXDLSTYMWIKFPRDEGACOPCPINCTHSCVDDLKCKCPAEORAS 649
DB 581 FQDQPCIPRCPHGMGLDGDTL-IMKXADMGGCQCHQNCCTQGGSPGLSGCRD-IVS 638
QY 650 PLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKXTMRRLQETELVEPLTPSGAMENO 709
DB 639 HSLIAGVLVSGLLITVALLIYVLLRRRRIK-RKXITIRLLQKEIIVEPLTPSGAANO 697
QY 710 AQMRILKEQYIKANSKFFIGTEL-TYKGIWLPDGNVNIPIAIVKLRNTPSPKANKEL 768
DB 698 AFIRILKETEFK-KDRLVSGAGFVYKGLMPPDGENIRIPAVIYVLEASIPKVNQVL 756
QY 769 DEAYVAVGSPYVSRLLGICLSTVQVLTQCLMPYAGCLLDHYRNGRLGSDILLNMQO 828
DB 757 DEAYVAVSDHPHVGRLIGLICLSAQVLTQCLMPYAGCLLDHYRNGRLGSDILLNMQO 816
QY 829 IAKGMYLDEVLYNHDLAARVLYKSPNHVKITTPGCLARLLIDIDETVYHADGGKVPILK 888
DB 817 IAKGMYLEERHLYNHDLAARVLYKSPNHVKITTPGCLARLLIDIDETVYHADGGKVPILK 876
QY 889 MALESILRRRFTHQSQVMSYGVTVWELMTFGAKPPDGIIPAREIPLLEKGRLLQPPICT 948
DB 877 MALESILQWYTHQSQVMSYGVTVWELMTFGSKPPYDGIIPAKIASVLENGBRLQPPICT 936

```

```

QY 949 IDVYIMYKCMWIDSECRPRRELVSERKARDPQRPVTVQNEDLAPASFLDSTFYRSL 1008
DB 937 IEVYIILKCMWIDSSRPRELVLGSEKQARQPSRYLVQ---NLPSFSDRLRSL 993
QY 1009 LEDDMGDLVDAEELVYVQGFPCPDPAAGVWHHRHSSSTRSGGDTLGLERSEE 1068
DB 994 LSSDD-DVDADELTL-----RIRKIN-RQGS----- 1018
QY 1069 EAPRSPPLAPSGAGSDVFGDLGMAKGLQSLPTHPSPLOYRSEDPTV-PLPSETDGY 1127
DB 1019 -----EPCIPNGH-----FVRENSIALRYISDPTQNALEKLDGH 1054
QY 1128 VAPLTCQSPRYVQPDVROP-----PSPR-----GPLP-AAPRAGATLERAKTLS 1174
DB 1055 -----EYVQPGSEYSSRLSDIYNRYEDLTGQGPVLSLQGEARTNFSREYLN 1104
QY 1175 PGRKGVNDVFAFGAVENPEYLPQCGAAPQHPHPAPAFADNLVYWDQDPPERGAP 1234
DB 1105 TNQNSL---PLVSSGSMDDPDY---QAG-----YQAF-----LPQTGALT 1139
QY 1235 STFKGTPTAENPEYLSL 1251
DB 1140 GNGMFLPALAENLYEYGL 1156

```

RESULT 5

```

ID 09M6F6 PRELIMINARY; PRT; 1137 AA.
AC 09M6F6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN BR34.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
PP SEQUENCE FROM N.A.
RC TISSUE=HINDBRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neurogranin-1 (nrg1) and erbb4 transcripts in
RT embryonic chick hindbrain.";
RL Mol. Cell. Neurosci. 13:237-258(1999).
DR HMBL; AF121963; AAD31764.1; -.
DR HSSP; P11362; 1FG4.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP_mocif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR Pfam; PF02757; YLP_2.
DR PRINTS; PRO0109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE, ATP, 1.
DR PROSITE; PS50011; PROTEIN KINASE, DOM, 1.
DR PROSITE; PS00109; PROTEIN KINASE, TYR, 1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW kinase; Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC4F CRC64;

```

Query Match 38.4%; Score 2615.5; DB 13; Length 1137;

Best Local Similarity 45.9%; Pred. No. 6,66-190;
Matches 524; Conservative 174; Mismatches 362; Indels 81; Gaps 28;

```

QY 161 LCYQDTLMKDI FHKNQALALTLITNSRACHPCSPMKCKSRWGSSSEDQGLRTVC 220
Db 3 LCFADITHMODIVRNPMASNFTLVPTNSSGGGRHKSCT-RWGPTENKCTLTITKVC 61
QY 221 AGGC-ARCKGPLPTDCHEOCAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDF 279
Db 62 AEOCDGRCYGVYVSOCHREGCGSGPKDTCFACNMFNDSGACVYQCTPFTYNTTF 121
QY 280 ESMNPBGRITFGACVYACPYNYLSTVSGCTLVCLHNEVTAEDGTORCEKSPCA 339
Db 122 QLEHNNAKYTYGACVYKCPNHFV-VDSSECVACSSKMEV-EENGIKKCKPCTDIP 179
QY 340 RVCYGLQMOKANSKFGITELF-PACCKKIFSLFLPESPFGDASNTAPLOPQLO 338
Db 180 KACDGIQGSU-VSAQYDSSNIDKFNCTKINGNLFLVTGIGHDYHTTAAINPEKLN 238
QY 339 VETLEITGVYISAMPDLSPLDSVFONLQVIRGILHNGAVSLTLQGLISMLGLSL 458
Db 239 IFQVREITGYLNTIOSWPEBNTDFRVFSLVLTIGRALYSGLSLILKQGITSLQFSL 298
QY 459 RELGGLALIHNNTHLCFVHTVPMDOLEFNPHQALLHTANRDECEGSELACHQCAR 518
Db 299 KOISAGNIYITDNSMLCYHTVNMSTLEFSTPSQKTVIHRKKKANCCTADMGVNECLSSD 358
QY 519 HGWGPTQCVNCSQFLRGOECVECEVYLOGLPREYVNAHCLPCHPECP-ONGSVTC 577
Db 359 GCMGPGDQCLSKCFIRGTCIESCNLYDEFEFANGVSCHECDQCKMEDNMITCY 418
QY 578 GPEADQVACAHYDPRFCVACSGYKPDLSYTPAKFPDEBACOPCINCTHSCVDL 637
Db 419 GPBDHCTCKHFHDGPRVCEKCDQGLGANSF--IKRYADEBCECHPCPHNCTOCGRP 476
QY 638 DDKG-----PAECRASPLTSIVSAV--GILLVAVLGVFGLIKRQOKIRXY 685
Db 477 ASHDCIYVPTWRTQSTLPQAR-TL--IAGVIGGLIYIMGLTFAVYRRKSIK-KKR 532
QY 686 TMRILQTEIVEPLTSGAMPNQAQRILKEOYIKANSKFGITEL-TYKGIWIPDE 744
Db 533 ALRFL-ETELVEPLTSGTAPNQAQRILKTELK-RVVLVLSGAFGYIKGIWIPDE 590
QY 745 NVKI PVAIKVLRNTSPKANKELIDEAYVAGVSPVSLIGICLSTVOLTQCMRPG 804
Db 591 TKIIPVAKIKINETGKRAVEFHEDELIMASDHPLVLLGVCLSPITQLTQMPHG 650
QY 805 CLDHYVNRGRGLSOLLMKQIAKMSYLEDLVLRDLAARVLYKSPHVKITDF 864
Db 651 CLDHYVHEKDNIGSOLLNWCQIAKMMYLEBRRLVHRDLAARVLYKSPHVKITDF 710
QY 865 GLARLLIDETEVYADGKVPIKMMALESTLRRRPTHOSVWSYGVTEMLTFGAKPD 924
Db 711 GLARLLDEDEYVADGKVPKIMALECHYRKFTHOSVWSYGVTEMLTFGAKPD 770
QY 925 GIPAREIPDLLEKGERLPPICTIDVYMIYKCMWIDSECRPREFRELSEFSRMAADPQ 984
Db 771 GIPAREIPDLLEKGERLPPICTIDVYMIYKCMWIDSECRPREFRELSEFSRMAADPQ 830
QY 985 RFVYIQMED-LGPASPLDSTFYRSLLEDMDMDVNAEELVYQGFPPDPAPAGGV 1043
Db 831 RYLVIQDDPKMLPSPDNSKFFQVLDEEDLEOMMAEEELVY-QAFNIPPIYTSRTRI 889
QY 1044 HHRHRSSTSGGGDLTLGLEPSEENAPRS--PLAP-SEGAGSDVDDGLGAKAGQS 1100
Db 890 DSNRQFVYRDGVAEEQGV-PMPYRAPGCIIEAVVAGATAEIEDCCNGTLRKQYA 948
QY 1101 LPTHDSPQLQYSEDPPVPLS-----ETDGYVAPLTCQPEYVNOVDVPPPSPR 1153
Db 949 TLAKEDSSTQYSDPVLPIPERVIGELDEDDGYMTPMRDKRTYLLNVEENPFVSRK 1008
QY 1154 EGRLPAA-RPAGATLERAKTLSPKNGVAVDF-----AFGAVENPEVLTQGGAA 1204

```

```

Db 1009 NCDLOAVNPEYHN-----APNGCPKADEYVNEPLVNTFANTLENAYL----- 1054
QY 1205 PQHPPEAFSAFDNLYYMDQDPERGA--PPSTFKGNPT-----AENPEY 1248
Db 1055 -KNLPERAKKAFNDPDMYHNSLSPRRSLQHDPVLQGSYTVFYKQNGRIPIVAENPEY 1113
QY 1249 L 1249
Db 1114 L 1114

```

```

RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
ID P79754;
AC P79754;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Brb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelline, K., Brenner S.;
RT Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes.
RL Genome Res. 9:251-258 (1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR0000494; EGFR_L domain.
DR InterPro; IPR000713; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KV ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

```

```

Query Match 32.6%; Score 2225; DB 13; Length 1328;
Best Local Similarity 39.6%; Pred. No. 4,3e-160;
Matches 511; Conservative 156; Mismatches 424; Indels 200; Gaps 34;

QY 9 WGLIALLP--GAAGN-----VCTGTDKMLLPSPFTHLMDLRLHYGCGVAGNLEI 62
Db 4 WRLIMVVARLAAASOTQOBAVCPETQGLSSTGQENQVYLNDRKXGCIINGNLEI 63
QY 63 TYLPNASTSLFDIOEGVGYLIANNQVRYPLQRLRVRGTOGFEDNYALAVDNDGP 122
Db 64 TOLESWVDFSLFTIRETVGYLIANNHFOELPLQGLRVIRNSLYERRFALSVLN--- 120
QY 123 LNNTPVYTAGSPGCLRELQRLSTELKXGVLIQRNPOLCYODTILMKDIFHKNNQAL 182
Db 121 ----YKDG--PSGLQLQELMNTLETLDGGVGI INNKYLRYPWYWDII--RNNDABE 173
QY 183 LIDNRSRACHPCSPMKSGSRWGSSSEDQGLRTVACAGC-ARCKGPLPTDCHEOCA 241
Db 174 IQNGERGVCH--KSC-GNYCMGKGXQCCQLTLITVCAPCQNDSCFSTSRDCHIECA 229
QY 242 AGGTGKSHDCLALCFHNSGICELHCPALVTYNTDFESMNPBGRITFGASCYTACPY 301
Db 230 AGCKGPLDTCFACRLFNDSGACVPCQDTLLYNNKQTFOMETNPNKQYQYSSICVSCPT 289

```

```

QY 302 NYLSTDVGSCTLVCPBLNQEVTADGTOR-CEKSKFCAQVVCYGLMQUYIKANSKFITG 360
DB 290 HFV-VDGSSCVSVCEPDMEV--ERGSQROCELOSGCPRYCEGTAE---QRQVSS 342
QY 361 ELE-FAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOQVFEETLEETGYLYISAMPDL 419
DB 343 NIDSFINCTKIOGSHFLVTGILGDDDFKNNPPLDAKLEFVRIRVRIITOLINIGMPREL 402
QY 420 PDLVFNQLQVIRGILHNGVSLITLQGLISWGLRLSELGSGIALIHNTLCEVHT 479
DB 403 NDLVFSLLITIOGSLFKRPSLWMMRIPTLISLGRSLREISDGSVYISONAHLCYHET 462
QY 480 VPMQDLFRNPH-QALLHTANRPEDECEGEGLAGCQCARHGCGPGPTOCVNCOSQPLRQ 538
DB 463 VMMTOLFRRGSRVRAINSNRPMAECVADGRVCDPLCSDCGCGPPDCLSCRYNSRIG 522
QY 539 ECVEECRVLOGIPREYVNAH-CLPCHPECOPONGSYTCGPADOCVAAHAKDPFCV 597
DB 523 TCVAGCHFNSGIPREFAGLVGVCACHPECKPOTGKASCCTGPADECMACTRFRDGPYCM 582
QY 598 ARCPGVPDLSTYMPIMKFPDEGACOPCINCTHSCVDDLDKGCAPAEGRASPLTSISA 657
DB 583 SECAPAGN-DEKGLIFKFPREHCEHCNCTOGSGSGGLND---LEAARLITSSQ 638
QY 658 VVGLLVVVLGVF-----GLIKRQOKIRKYMRLQETELVEPLUTSGAMPQA 710
DB 639 ITGIALGVPAGLIFCLVLFPLGMLYHGLAIRRKAMRRYTESGESFEPLGP-GEKGTXY 697
QY 711 QMRILKEQYK---ANSKFIGITELTVYKGIWIPDENVKIIVALKVIRENTSPKANE 766
DB 698 HARILKPSDLKIKPLSGVFG---TVSKGFMPIBEETKIVALKTIQDSGROTFE 753
QY 767 ILDEAYVAVGVSFVSESLGICLTSTVQVLTQIMPYGCLLDHVENREGLSGDILLWC 826
DB 754 ITDHLISWGSLDHPRIYVLLIGICGTCICLVLTQLSHSGSLHEIRHOKTSLDPORLIWC 813
QY 827 MOIAKMSYLEDVRLYHMDLAARVNLKSPHNKXITPFGARLLIDETEHADGKVP 886
DB 814 VOIAKMSYLEDHVKHNLAAKNIILKNDYQVQISYVADLLYDDKIVYSEKTFPI 873
QY 887 KMALESTILRRFTHOSDVSQVTVVWELMTFGAKPYDGIIPAREIDLLKEKRLPQPI 946
DB 874 KMALESTILFRYTHOSDVSQVTVVWELMTFGAKPYDGIIPAREIDLLKEKRLPQPI 933
QY 947 CTIDVYIMVYKCMWIDSECRPRELSEFSRMAPRQRFVVIQNDLGPASPLDSTFR 1006
DB 934 CTIDVYIMVYKCMWIDSECRPRELSEFSRMAPRQRFVVIQNDLGPASPLDSTFR 980
QY 1007 SLEDDDMGDLVDAEYVLVPOQGFCCPPAPGAGGVNHHRRSSSTRSGGDLTGLEPS 1066
DB 981 --EDSGMGBFL-----RGSER---GLLEADLEED 1005
QY 1067 EEEARSPSLAPSEGSQVDFDGLMG--AAKGLOSLPHDPSPLQ-----RYS 1113
DB 1006 EEE-----GLDPRATSLQSPSWMSTSPQINSYVMMTQLRND 1044
QY 1114 EDPVPLPSETDGVAPLTGSPQ-ERYNQ-----PDVRQPPSPREGP- 1157
DB 1045 ----FAVSOQGIIGLYPMSFVDITRQWYQRLSRTVTLPRSAFRSSREAEJC 1098
QY 1158 -PAARPAAGATLERAKTISPCKNGVAVKVFAGAVNPELTPQGGAAQHPHAPAFSPA 1216
DB 1099 EDGAGCAGIFRVR-----FSGERKN-----PQGG----- 1122
QY 1217 PDNLVYWDQPPERGAPSTFKGTPTAENPE 1247
DB 1123 -----QQRKLSTASSPSSFKTWADEDE 1146

```

RESULT 7
098I9
ID 098I9 PRELIMINARY; PRT; 1433 AA.
AC 098I9;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
OS
GN Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR Interpro; IPR000345; CyC heme bind.
DR Interpro; IPR000484; EGFR_L_domain.
DR Interpro; IPR000719; Euk_kinase.
DR Interpro; IPR002174; Furin-like.
DR Interpro; IPR002290; Ser_thr_kinase.
DR Interpro; IPR01245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00693; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PRO0109; Tyrosine_kinase.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; F0; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYKC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 28.9%; Score 1971.5; DB 5; Length 1433;
Best Local Similarity 32.1%; Pred. No. 9,4e-141;
Matches 465; Conservative 198; Mismatches 399; Indels 387; Gaps 39;
QY 26 CCGTDMKRLRPSFPTLMDRLHLQSGCVVQSLLELYFTNASLFLDDIQGVGYVL 85
DB 1 CIGTGRMSVPMANREYHYNLRDRYTNCTVYDGNLEITWIONITDLNFIHREVTGYVL 60
QY 86 IAHNVQVPLQRLRIVGTOLF-----EDNYALAVLDNGPLNNTTPTVGSPGGLREL 140
DB 61 ISLYDLRYVILRLRIIGRTTFKLKNEEAYGLFV-----SFSMNTL 104
QY 141 QRLSTEILKGVLIQRPOLCYODTIMKDI-FHKNQALALITDNRSRACHPGSPMC 199
DB 105 ELPALRLDIGSGVGFNNYNLCHMSIMWEILLAPOTSQYTFNFSSPRVCPCHPSC 164
QY 200 KSGRCWSESSQCSLTTCVACGGA--RCGRLPTDCCHEGCAAGCTGKSHDCLACH 257
DB 165 EYG-CWGBAHNCQFSLKNSPQSGRCGPRPREDCHFCAGGCTGTGSDCLACKN 223
QY 258 FHNSGICELHCPALVYNTDTFESNPNEGRYTFGASCYTAAPNYLSTDVSGCTLVCP 317
DB 224 FYDDGVCKQECPPMQIYVPTVYFBNPDGKXAGATCVAKCP-ENHLKNGACVAKCPK 282
QY 318 HNGEYTAEDGTORCKSKGPCARVCYVGMGYIKANSKFIGITELBEPAGCKTIFGSLAF 377
DB 283 GKMPQNSE-----CYPCGVCPKTCGEGE---VHSDNIG---NYKDTTIEGSLLE 329
QY 378 PESFDGDPASNT-----APLOPELOQVFEETLEETGYLYISAMPDLSPDLVFNQ 429
DB 330 DQSFQFQCVYNTFESFGPRYKIDPDRLEVFSTVKEITGFINIQAHNPFTTLNYPNLE 389

```

```

QY 430 VIRGRILHNGAY-STLLOGLISWLGRLSRLBLSGLALIHNNHLCFVHTVMDOLFNR 488
DB 390 VVGGRQLKENLFASYIYIKTSIKSLKSLKRVNSGIVILENSDLCFVEDIDSEITKS 449
QY 489 PHOALLHTANRPEDCEVGBGLAQHLCARGHCWGPPTQVCNCSQFLRGCEVECECVLO 548
DB 450 SDHEVWVKRNATCHEHEGMECESECSKACWGWGPGQCECNVYKXKGLDSCCK--- 506
QY 549 GLPREY-VNARHCLCHPECCORONGSVTCFGRPADOCACAHYDPPFCARCG----- 601
DB 507 SLPLYSVDSKTCGCHQCECKD----FCYGNBDCGSCNNVXDRGFCVAECPTTHAM 561
QY 602 -----SGVKPDLSTYPIWKFPD----- 618
DB 562 NGTCINCHKTCVCGCRPRDTIAPDCISCDKAILIGSDAKIERCLMKDESCPDGYSDYVL 621
QY 619 -BEG----- 621
DB 622 QEEGPLKOLSGKAVRKCHPRCKCTGTGYGHEPFCQECTGYKGECEDECPDQFYANEE 681
QY 622 --ACQPCPINCT-----HSCVDL-----D-----KCGPABQ----- 646
DB 682 TRICLPHQECRGCHGLDGDHHECENLKLFEGRDYDNATFTCVSNCRPAHPRKRPQEA 741
QY 647 -----RASPLTISAVNGILLVVLGVVFGI---LIKROQKIRKRYTM 687
DB 742 GRIGPYCSADSMOSGLRIEPOQYKIVWGSYMALILLCVFGLAFVLFSSHKKKXAVVM 801
QY 688 RRLLOETELVEPLTPSGAMPQAOIRIKEGYIANGSKFGLTEL---TVYGGIMIPGENV 746
DB 802 TMLAGCEDSEPLRSNVGNPLTKLRIIKEKEIRGG-VLGMGAFGVFGVMMPEGESV 860
QY 747 KIPVAIKVLRNTSPKANKELLDEAYVWAGVSPYVSRLLGICLSTVQVLVTOIMPYGCT 806
DB 861 KIPVAIKVLMEMSGESSEKFELEAYIMASVEHBNLLKLAVCMTSGMMLITOLMLGCI 920
QY 807 LDHVENRGRGLSQDILNMCQIAKMSYLEDVLYVRDLAARVLYKSNHYKXITDFGI 866
DB 921 LDYVANNKOKTIGSKALLWSTQIARGMAYLEERLVHRDLAARVLYVOTSSCVKXITVFGI 980
QY 867 ARLLDIDETEYHADOGKVPKIMMALESILRRFTHOSDVSYGVTWELMTFPAKPYDGI 926
DB 981 AKLLDFDSEYRAAGKMPKIMALECIRHRVFPSKSDVWAFGITIMELLITGARYENV 1040
QY 927 PAREIPDLLEKGEELPQPTICTIVYKIMYKCMITDECPRRRELVSFGRMADPQGF 986
DB 1041 PAKDVPLEIEIGHKLPQPDICSLDYCILLSCWVLADARPTFKOLAETFAEKARDPGHY 1100
QY 987 VVIONEDLGPASPLDSTFYRSLIEDDMGDLV----- 1018
DB 1101 LMI-----PQDKRMRLPSYTNODEKDLIRTLAVVAMAAAAAAGASVNDVPSTI 1151
QY 1019 -DAEYLVPOQGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEAPRS--- 1073
DB 1152 AETDEYLOPKTRPSIMLPBSA-----VEPS-DEMPKSLRYC 1187
QY 1074 --PLAP---SEGASDVFDDDLGWAAGKGLSLPTHDPSPLCYSSDPVPLPSETDGYV 1128
DB 1188 KDPILKPDDETIDGHKEV-----GVGGIR-----LMLPLDDEDVL 1221
QY 1129 APLTCSPOPEYVNPQVPRPSPRREGPLPABPAGATERAKTLSPGKXGVKVFARFG 1188
DB 1222 MF-TCOSQ---NQS-----TPG---YMDLIGVP 1242
QY 1189 GAVENPEYL-----TPGGGAAPQHPBPAPAFSPAFNLYWDDDPERGAPSTFGX 1239
DB 1243 ASVDNPEYLMOSTAATAGLAQSGWG-FHTPP-----PPTWNG 1279
QY 1240 TPTAENPE 1247
DB 1280 MPTHQHSQ 1287

```

```

RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Herectatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RP [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF17761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FFCIBEE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 96.8%; Pred. No. 6.6e-134;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALLPFGAASSTOVTGTDMLRLPASPETHLMLRHYOGCQVQGNL 60
DB 1 MELALCRWGLLALLPFGAASSTOVTGTDMLRLPASPETHLMLRHYOGCQVQGNL 60
QY 61 ELYTLPNASSLFTQDIQEVQGYVLAHNOVROVPLQRLRYVGTOLFEDNALAVLDNG 120
DB 61 ELYTLPNASSLFTQDIQEVQGYVLAHNOVROVPLQRLRYVGTOLFEDNALAVLDNG 120
QY 121 DPLNNTTPVGSAGSGGRLREQLSLTEILKGYLIRNPQLCYQDITILKDIFFHXNQLA 180
DB 121 DPLNNTTPVGSAGSGGRLREQLSLTEILKGYLIRNPQLCYQDITILKDIFFHXNQLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGRCKWGSSESDCQSLRTVYACAGGACRCKGPLPTDCHEOC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCKWGSSESDCQSLRTVYACAGGACRCKGPLPTDCHEOC 240
QY 241 AAGCTGKHSDDCLACHFNHSGICEHCPALVYVNTDRESMNPBGRITTFGASCYTACP 300
DB 241 AAGCTGKHSDDCLACHFNHSGICEHCPALVYVNTDRESMNPBGRITTFGASCYTACP 300
QY 301 YNYLSTDVSGCTLVCPLNQEVYAEADGTORCEKCSKPCARVCYGL 345
DB 301 YNYLSTDVSGCTLVCPLNQEVYAEADGTORCEKCSKPCARVCYGL 345

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RA Scienceberg R.
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0B503395F9E101B0 CRC64;
 Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 6e-124;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;
 QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIDLLKEGRLLPQPICT 948
 DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIDLLKEGRLLPQPICT 60
 QY 949 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRL 1008
 DB 61 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRL 120
 QY 1009 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSQPEEVYNQPDVPRPQPSREGPLPAAPAGATLERAKTLPFGKGVKVDVAFG 1188
 DB 241 APLTCSQPEEVYNQPDVPRPQPSREGPLPAAPAGATLERAKTLPFGKGVKVDVAFG 300
 QY 1189 GAVENPEYLTPOGGAAPQPPPAFNDLYWDDPBERGAPPTFGCTTAENPEY 1248
 DB 301 GAVENPEYLTPOGGAAPQPPPAFNDLYWDDPBERGAPPTFGCTTAENPEY 360
 QY 1249 LGLDVTV 1255
 DB 361 LGLDVTV 367
 RESULT 10
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN P3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Theria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.,
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF18349; AA5586.1; -
 DR InterPro; IPR002048; E-hand.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;
 QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIDLLKEGRLLPQPICT 948
 DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIDLLKEGRLLPQPICT 60
 QY 949 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRL 1008
 DB 61 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRL 120
 QY 1009 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSQPEEVYNQPDVPRPQPSREGPLPAAPAGATLERAKTLPFGKGVKVDVAFG 1188
 DB 241 APLTCSQPEEVYNQPDVPRPQPSREGPLPAAPAGATLERAKTLPFGKGVKVDVAFG 300
 QY 1189 GAVENPEYLTPOGGAAPQPPPAFNDLYWDDPBERGAPPTFGCTTAENPEY 1248
 DB 301 GAVENPEYLTPOGGAAPQPPPAFNDLYWDDPBERGAPPTFGCTTAENPEY 360
 QY 1221 YWPD-QDPPER-----GAPSTFGTPTAEN 1245
 DB 361 YWPD-QDPPER-----GAPSTFGTPTAEN 410
 RESULT 11
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviruses; Retroviridae; Alpharetrovirus.
 CX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynosccheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Berg H.,
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02813; Retro M; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84DF6514FE1D63 CRC64;
 Query Match 24.1%; Score 1640; DB 15; Length 729;
 Best Local Similarity 53.2%; Pred. No. 5.7e-116;

Matches 348; Conservative 75; Mismatches 127; Indels 104; Gaps 17;

QY 569 PONGSVTCGPEADCVACAHYKDPFCVACPSGVKNDLSYMP1WKPPEBEGACGCPPI 628
 DB 141 PEETATPKTG--DHCKKCAHFIDGPPCVACCPAGVIGENDTL-VKRYADANAACQLCHP 197
 QY 629 NCTHSCVDLDKGCPCAEORASPLTIVSAVY-GILLVVLGVVFGILIKRQCKIRKRYTM 687
 DB 198 NCTRGCKGFGLEGCP---NGSKTPIAAGVVGILLCLVVGILGIVLRRR-HIVRRTLT 253

QY 688 RLLOETELVEPLTSSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGDEVN 746
 DB 254 RLLOERELVEPLTSSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGDEVN 746
 QY 747 KIPVAIKVLRNTSPKANKELIDEAYVAGVSPYVSRLLGICTSTVOLTPMYPGCL 806
 DB 313 KIPVAIKELRENTSPKANKELIDEAYVAGVSPYVSRLLGICTSTVOLTPMYPGCL 806
 QY 807 LDHVRNRRGLSGODLWKCQIAGKMSYLEDVRLVHRDLAARNVLKSPNHYKITDPGL 866
 DB 373 LDYIREHKDNISQYLLWMCVQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDPGL 432

QY 867 ARLLDIDETEHADGKVPYIKMALESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI 926
 DB 433 AKLLGADKEVHAEGKVPYIKMALESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI 926
 QY 927 PAREIPDLLEKGERLPQPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQR 986
 DB 493 PASEISSVLEKGERLPQPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQR 986

QY 987 VVIO-NEDLGPASPLDSTFYRSLLEDMDGLVAEEVLPQGFPCPDPAFGAGMWH 1045
 DB 553 LVIOGGERMHLPSPTDSKFRYRLTMEEDMEDIVDAEYLVPHQGF----- 598

QY 1046 RHRSSTRSGGDLTLGLEPSEEBAPRSPL-----APSEGAGSDVPDGLGMGAKGLQS 1100
 DB 599 --NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQG 630

QY 1101 LPTHDPSPLOKSEPTVPLPSET--DGVAAPLTGSPQPEYVNPDPVPQPSPREGFLP 1158
 DB 631 HVRREDSFVQRYSSDPTNLFLESIDDGFL-----PAPEYVNO--LMPKFS----- 675

QY 1159 AARPAGATLERAKTLPGRNGVVKDF-----AFGAVENPEYL 1197
 DB 676 -----TAMVQNOYNNISLTAISKLPMDSRKYNSHSTAVDNPETL 715

RESULT 12
 Q86714 PRELIMINARY; PRT; 567 AA.

AC 086714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OC NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; .
 DR HSP; F11362; 1FGK.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; kinase; 1.

DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; TykKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25EL CRC64;

Query Match 24.0%; Score 1638; DB 15; Length 567;
 Best Local Similarity 53.8%; Pred. No. 5.5e-116;
 Matches 347; Conservative 73; Mismatches 121; Indels 104; Gaps 17;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKNDLSYMP1WKPPEBEGACGCPPI 637
 DB 1 GP--DHCKKCAHFIDGPPCVACCPAGVIGENDTL-VKRYADANAACQLCHPCTRGCKGP 57

QY 638 DDKGPAEORASPLTIVSAVY-GILLVVLGVVFGILIKRQCKIRKRYTMRLLOETEL 696
 DB 58 GLEGCP---NSKTPSIAAGVVGILLCLVVGILGIVLRRR-HIVRRTLRLLQEREL 113

QY 697 VEPLTPSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGDEVN 755
 DB 114 VEPLTPSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGDEVN 755

QY 756 RENTSPKANKELIDEAYVAGVSPYVSRLLGICTSTVOLTPMYPGCLLDHVRNRRG 815
 DB 173 RENTSPKANKELIDEAYVAGVSPYVSRLLGICTSTVOLTPMYPGCLLDHVRNRRG 815

QY 816 RIGSODLWKCQIAGKMSYLEDVRLVHRDLAARNVLKSPNHYKITDPGLARLLDDET 875
 DB 233 NIGSODLWKCQIAGKMSYLEDVRLVHRDLAARNVLKTPQHVKITDPGLARLLDGADEK 292

QY 876 EYHADGKVPYIKMALESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGPAREIPDL 935
 DB 293 EYHADGKVPYIKMALESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGPAREIPDL 935

QY 936 EKGERLPQPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQRFVVO-NEDL 994
 DB 353 EKGERLPQPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQRFVVO-NEDL 994

QY 995 GPASLDSTFYRSLLEDMDGLVAEEVLPQGFPCPDPAFGAGMWHRRSSSTRS 1054
 DB 413 HLPSTDSKFRYRLTMEEDMEDIVDAEYLVPHQGF-----NSPST-- 454

QY 1055 GGGDLTLGLEPSEEBAPRSPL-----APSEGAGSDVPDGLGMGAKGLQSLPTHDPSP 1109
 DB 455 -----SRTPLLSLSATSNNSATNCID-----RNGGHVREDSFV 490

QY 1110 QRYSDPTVPLPSET--DGVAAPLTGSPQPEYVNPDPVPQPSPREGFLPAARPAGATL 1167
 DB 491 QRYSDPTGPNLFLESIDDGFL-----PAPEYVNO--LMPKFS----- 526

QY 1168 ERAKTLSPGKNGVVKDF-----AFGAVENPEYL 1197
 DB 527 -----TAMVQNOYNNISLTAISKLPMDSRKYNSHSTAVDNPETL 566

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.

AC 064895;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gag; V-erb-A; V-erb-B protein.
 GN GAG; V-ERB-A; V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE:90206603; PubMed:1969616;
 RA Brustin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B.";
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL: X52309; CA36459.1; -.
 DR HSP; P10828; 2NL.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000536; Hormone_rec_119.
 DR InterPro: IPR001723; Stehnm_receptor.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR001628; ZnF_Csteroi.
 DR Pfam: PF00104; hormone_rec_1.
 DR Pfam: PF00069; kinase_1.
 DR PRINTS: PR00398; STRDOMONER.
 DR PRINTS: PR00047; STRODINGER.
 DR PRODOM: PD000001; Euk_pkinase_1.
 DR PRODOM: PD000035; ZnF_Csteroi; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR SMART; SM00399; ZNF_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AD791E4E95CE CRC64;
 Query Match 23.2%; Score 1581.5; DB 15; Length 962;
 Best Local Similarity 50.4%; Pred. No. 2.4e-111;
 Matches 349; Conservative 73; Mismatches 150; Indels 121; Gaps 20;
 QY 541 VEEGRVLOGLPRE-VYNNR-HCLP-----CHPEQC 568
 DB 354 IEKCESTYLLAFEHINRKNNITHFMSKILMKVADLMIGAYHASFELMKVECPTELS 413
 QY 569 PQNGSVTCFGPEADQCVACAHYKDPFCVACPSGVCXPDLSYMPKFPDEGACQPCPI 628
 DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHP 465
 QY 629 NCTHSCVLDKCGCPAEGRASPLTSISAVV-GILLVVLGVVFGIILKRRQKIRKRYTM 687
 DB 466 NCTRGCKGPGLEGCP--NGSKTPSIAAGVGGLLCLVWGLGIGLYLRRR-HIVKRTLL 521
 QY 688 RRLQETELVEPLTPSGAMPNQAOMRIKEQYIKANSKFTIGTEL-TVYKGIWIPGGENV 746
 DB 522 RRLQERLVEPLTPSGARPQALRIILKEFERK-KYKVLGFGAFGVTVYGLMIPGEKV 580
 QY 747 KIPVAIKYLRNTPSPKANKELLDEAYVAVAGVSPYRLLGICLTSTVOLVTLQMPYGC 806
 DB 581 TIPVAIKELREATSPKANKELLDEAYVAVASVDNPHVCRLLGICLTSTVOLVTLQMPYGC 640
 QY 807 LDHYRENNGRIGSODLLNMCQIAKGSYLEDVRLVHRDLAARNVLYKSNHVKITDFGL 866
 DB 641 LDYIREHKONIGSOYLLNMCVQIAKGMVYLERHMHVARDLAARNVLYKTPQHKITDFGL 700
 QY 867 ARLLIDETEHYHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 926
 DB 701 AKQLGADKEHYHAGGKVPKIKMALESILHRYTHQSDVMSYGVTVWELMTFGSKRYDGI 760
 QY 927 PAREIPDLLEKEGELPQPCTITIVVMIMKCMWIDSECPRRRELVSFSGMARPPQPF 986
 DB 761 PASEISSVLEKEGELPPQPCTITIVVMIMKCMWISGASDPKRELIAPFSKXARDPPEY 820
 QY 987 VVIO-NEDLGSPASFLDSTFYRSLDEDDMDGLVDAAEYLVPOQGFCCPDPAPGAGMVMH 1045
 DB 821 LVIOGDERMHLPSPTDCKFYRTLMEEEDMEDIVDADYLVPHQGF----- 866

QY 1046 RHRSSTRGSGDUTLGLPEPEEAPRSPPLAPSEGAGSDVFDGLGMAAGLQSLPTH 1105
 DB 867 --NSPST-----SRPLPSSLSASIN-----NSATKICIRNGH- 898
 QY 1106 PSPFLORYSEDPVLPSPSTGDIYAPLTCSPQPEYVNPDPVAPQPSREGPLPAPAPAGA 1165
 DB 899 -----PVEDGFL-----PAPEVNO--LMPKRPSTAWQNIYNYISL 935
 QY 1166 T-LEBAKTLSPKNGVVDVFAFGAVENPEYL 1197
 DB 936 TAISKLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
 Q85468 PRELIMINARY; PRT; 545 AA.
 ID Q85468
 NC Q85468;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian Erythroblastosis virus (T334) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_Taxid=11861;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Hayman M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL; X06943; CA30024.1; -.
 DR HSP; P11362; IFCG.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; kinase_1.
 DR PRODOM: PD000001; Euk_pkinase_1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DCE8C0A0F8A74 CRC64;
 Query Match 23.1%; Score 1573; DB 15; Length 545;
 Best Local Similarity 53.4%; Pred. No. 4.6e-111;
 Matches 336; Conservative 70; Mismatches 129; Indels 94; Gaps 17;
 QY 578 GPEADQCVACAHYKDPFCVACPSGVCXPDLSYMPKFPDEGACQPCPI NCTHSCVDL 637
 DB 1 GP--DHCKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHP NCTRGCKGP 57
 QY 638 DDKCPAEGRASPLTSISAVV-GILLVVLGVVFGIILKRRQKIRKRYTM RRLQETEL 696
 DB 58 GLEGCP--NSKTPSIAAGVGGLLCLVWGLGIGLYLRR-HIVKRTLL RRLQERL 113
 QY 697 VEPITPSGAMPNQAOMRIKEQYIKANSKFTIGTEL-TVYKGIWIPGGENV KIPVAIK 755
 DB 114 VEPITPSGAMPNQAOMRIILKEFERK-KYKVLGFGAFGVTVYGLMIPGEKV TIPVAIK 172
 QY 756 RENTSPKANKELLDEAYVAVAGVSPYRLLGICLTSTVOLVTLQMPYGLDHYRENNGR 815
 DB 173 REATSPKANKELLDEAYVAVASVDNPHVCRLLGICLTSTVOLVTLQMPYGLDHYRENN 232
 QY 816 RLGSQDILLNMCQIAKGSYLEDVRLVHRDLAARNVLYKSNHVKITDFGLARLLIDET 875
 DB 233 NIGSOYLLNMCVQIAKGMVYLERHMHVARDLAARNVLYKTPQHKITDFGLAKQLGADK 292
 QY 876 EYHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDL 935
 DB 293 EYHAEKGKVPKIKMALESILHRYTHQSDVMSYGVTVWELMTFGSKRYDGI PASEISSVL 352

```

QY 936 EKERLPPICTIDVYIMVCMWIDSECRPRFRELSEFSRMARDQRFVIG-NEDL 994
DB 353 EKERLPPICTIDVYIMVCMWIDSECRPRFRELSEFSRMARDQRFVIG-NEDL 412
QY 995 GPASPLDSTFFRSLEDDMGDLVAEEYLVQCGFCPPDPAAGVGHHRSSSTRS 1054
DB 413 HLPSPDSTFFRSLEDDMGDLVAEEYLVQCGFCPPDPAAGVGHHRSSSTRS 454
QY 1055 GGGDLTLGLEPSEEAERSP-----APSEGAAGSVFQGDLCMGAKGLQSLPTHPDSPL 1109
DB 455 -----SRTPLLSSLSATSNNSATNCTIRNG-----H----- 481
QY 1110 QRYSEDPVPLPSEBDGYVAPLTCPPQPEYVQPDVRQPPSPREGPLPAAPAGAT-LE 1168
DB 482 -----PYREBDGFL-----PAPEYVQ-----LMPKPSSTAMVQIQIYNYISLTAIS 522
QY 1169 RAKTSPGKNGVVKVDFAGAVENPEYL 1197
DB 523 KLPMDSRVQV-----SHSTAVDNPETL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Scheil C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maithe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maithe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt P., Suzuki R., Tomita M., Wagner U., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barin G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wymshae-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayaishiaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275366; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 22.1%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 7e-106;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

```

```

QY 11 LTLALPFGAA--STVCTGDMKRLPASPETHLDMRLHYQCGVQGNLELTYPTN 68
DB 14 LTLALCAAGALAEKVKQGSNLTGCTGDEHFLSLQRYNNCEVVLGNLETTYQGN 73
QY 69 ASLSTLQDIQEVQGVYLIANQVROVLRIRIYRGTLQFEDNYALALVDNGDPLNTP 128
DB 74 YDLSTLKTIOEVAGVLLIALTNTVRIPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGRLRELSTLEIKGGLVLIQNPQLCYOPTILMKDI-----FKNNQLATLI 184
DB 125 -YGTNRTGRLPELPMENLEILIGAVRSNPILCNMTTIQKRDIVQVNFSSNMDL--- 180
QY 185 DTRSRACHPSPCMKGSRCWGESSEDDQSLTRTVCAAGCA-RCKGFLPTDCHEQCAAG 243
DB 181 -QSHPSGCPKCDPSCPNSSCWGGSENGCKLTKIICAQGSHRCGRSPSCGHNQCAAG 239
QY 244 CTGPRGSCCLACLNHNSGTELELCPALVTYNTTFSSMPREPRRTYFGASCYACGYNY 303
DB 240 CTGPRSCCLVQKQFQDQATKCTCPPLMLNPTTYQMDVNPBEGKVSFGATCVKCKPRNY 299
QY 304 ISTDVGSCTLVCPILHNOEVTAEQDGTQRCCKSCRCARVCYGLGQVYKASKFIGITELE 363
DB 300 VVTDHGSQVRAAGCPVYEV-BEDGIRCKCKDGCRCRVACNGIGIGERK-DTSLNATNIK 357
QY 364 -FAGCKKTFGSLATLPESFDGDPASNTAPLOPBOLOVFTLEETGYLYISAMPDSLPL 422
DB 358 HFKYCTALSGDLHLPLVAFKDSFTTRPPDPRLELTKVKEITGFLLIQAMPDMNTDL 417
QY 423 SVFQNTLVYIRGRILHNGAVSLTLQGLGSIWGLRSLELSGLALIHNNHLPFVHTVPM 482
DB 418 HAFENLEIIRGTRKHQGFSLAVGLNITSGLRSLEISDGVITISGNKNLTANTYIM 477
QY 483 DQLFNPQALLHTANREDECVGEGLACHQLCARGHGCMWPGPTQVNCQSOLPRGECVE 542
DB 478 KTLFETPQKTKIMNPRLEKCKKLVNHVNCPLCSSGSCWGPBEPDVCQNVSRGRECE 537
QY 543 ECRVLYQGLPREYVYARHCLPCHRECOQONSVCYCFGEADQCVACAHYKRPCTVACRPS 602
DB 538 KNILGEPREFVENSECTIQCHPCLDQAMNITITGSGPNCTQCAHYIDGPHCVKTCPA 597
QY 603 GVKPDLSTYPMKFPDEGACQCPINCTHSCVDLDRKGC 642
DB 598 GIMGENNTL-VKXADANNVCHLCHACTGCGAGPGIQC 636

```

Search completed: July 22, 2003, 09:24:34
 Job time: 51.8246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.9339 Seconds

(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 1 MELALACRWGLLALLPPGA.....TFKGTPTNPEYLGDVPR 1255

Sequence: 1 MELALACRWGLLALLPPGA.....TFKGTPTNPEYLGDVPR 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSz/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	6646	97.5	1255	21 AAY92620 Human heregulin 2
2	6646	97.5	1255	22 AAE12130 Human tyrosine kin
3	6646	97.5	1255	22 AAB60167 HER2 transgene pla
4	6646	97.5	1255	23 AAU74545 Human HER2 (ErbB2)
5	6640	97.4	1255	17 AAU01111 HER-2/neu protein.
6	6640	97.4	1255	20 AAB22466 Human HER-2/neu on
7	6640	97.4	1255	21 AAB21198 Human HER-2/neu pr
8	6640	97.4	1255	21 AAB21198 Amino acid sequenc
9	6640	97.4	1255	22 AAB85458 Human HER-2/neu pr
10	6640	97.4	1255	22 AAG88267 HER2/neu amino aci

11	6640	97.4	1255	23 AAB24067 Human Her-2 protei
12	6640	97.4	1255	23 AAE20479 Human Her-2/neu pr
13	6640	97.4	1255	23 AAM51143 Human Her-2/neu on
14	6640	97.4	1255	24 AAU77114 Human Her-2/neu po
15	6597	96.8	1433	14 AAB39568 Sequence of c-erbB
16	6476	95.0	1223	23 AAU98923 Human breast cance
17	6323	92.8	1200	21 AAB21208 Human HER-2/neu pr
18	5853.5	85.9	1256	21 AAB21199 Rat HER-2/neu prot
19	5853.5	85.9	1256	23 AAM51144 Rat HER-2/neu onco
20	5829.5	85.5	1256	21 AAB21206 Mouse Her-2/neu pr
21	5829.5	85.5	1256	22 AAG2860 Amino acid sequenc
22	5829.5	85.5	1256	23 AAM51151 Mouse Her-2/neu on
23	4814	70.6	919	21 AAB21203 Human HER-2/neu fu
24	4814	70.6	919	23 AAM51148 Her-2/neu extracel
25	4064.5	59.6	920	23 AAM51152 Mouse Her-2/neu ex
26	4064.5	59.6	926	23 AAM51153 Mouse Her-2/neu ex
27	3698	54.3	712	21 AAB21204 Human HER-2/neu fu
28	3698	54.3	712	23 AAM51149 Her-2/neu extracel
29	3552	52.1	782	18 AAM19764 Her-2/neu CSF immuno
30	3550	52.1	653	21 AAB21200 Extracellular HER-
31	3550	52.1	653	23 AAM51145 Human Her-2/neu on
32	3512	51.5	645	22 AAB60408 Human ErbB2 oncopr
33	3512	51.5	645	22 AAB61593 Human ErbB2 extrac
34	3447	50.6	951	21 AAY44993 DC8acFv-erbB2EC fu
35	3344	49.1	624	11 AAR08222 Extracellular port
36	3067	45.0	1210	21 AAB39259 Amino acid sequenc
37	3067	45.0	1210	21 AAY50616 Human EGF receptor
38	3067	45.0	1210	23 AAE23019 Human Her-1 protei
39	3067	45.0	1210	23 AAM50768 Human epidermal gr
40	3065	45.0	1210	22 AAB68440 Amino acid sequenc
41	3053.5	44.8	654	21 AAB21205 Rat HER-2/neu prot
42	3053.5	44.8	654	23 AAB21150 Rat HER-2/neu onco
43	3026	44.4	1210	23 AAB21768 Human epidermal gr
44	2996	44.0	583	23 AAE20453 Human protein for
45	2996	44.0	587	23 AAE20481 Human protein for

ALIGNMENTS

RESULT 1	AA92620	standard; Protein: 1255 AA.
ID	AA92620	
AC	AA92620	
XX	10-AUG-2000	(first entry)
DT	10-AUG-2000	
XX	Human heregulin 2 (Her2).	
DE	Human heregulin 2 (Her2).	
XX	Her2: Vaccination: cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= "N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT /label= Cysteine_rich_domain
 FT Region 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 FT /label= ligand_binding_domain
 FT Region 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 359..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 455..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 FT /label= Cysteine_rich_domain
 FT Region 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 FT /label= Transmembrane_domain
 FT Region 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 FT /label= Tyrosine_kinase_domain
 FT Region 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 FT /label= C-terminal_domain
 XX
 PN WC200020027-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 05-OCT-1999; 99MO-DK00525.
 XX
 PR 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX
 PA (MEBI-) M & E BIOTECH AS.
 XX
 PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G,
 XX
 DR WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198, 220PP; English.
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animal immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6646; DB 21; Length 1255;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2;

QY 1 MELALICRMGLLALLPFGAASVCTGDMKRLRPASPTHLDVRLHYGGCCVQGNL 60
 DB 1 MELALICRMGLLALLPFGAASVCTGDMKRLRPASPTHLDVRLHYGGCCVQGNL 60
 QY 61 ELTYLPTNASLFLDDIOEVQGVLIANOVQVPLQRLIRVGTQLEFDNYALAVLDNG 120
 DB 61 ELTYLPTNASLFLDDIOEVQGVLIANOVQVPLQRLIRVGTQLEFDNYALAVLDNG 120
 QY 121 DPLNNTPTPTGASPGGLREQLQRLSTELTKGGVLIQRNPOLCYQDTILMKQIFPKNOLA 180
 DB 121 DPLNNTPTPTGASPGGLREQLQRLSTELTKGGVLIQRNPOLCYQDTILMKQIFPKNOLA 180
 QY 122 DPLNNTPTPTGASPGGLREQLQRLSTELTKGGVLIQRNPOLCYQDTILMKQIFPKNOLA 180
 DB 122 DPLNNTPTPTGASPGGLREQLQRLSTELTKGGVLIQRNPOLCYQDTILMKQIFPKNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCWSESSDDQSILTRFVAGAGCARCGPLPTDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWSESSDDQSILTRFVAGAGCARCGPLPTDCHEQC 240
 QY 181 LTLIDNRSRACHPCSPMKSGRCWSESSDDQSILTRFVAGAGCARCGPLPTDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWSESSDDQSILTRFVAGAGCARCGPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELKCPALVTYNTDFESMPNDEGYTTGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELKCPALVTYNTDFESMPNDEGYTTGASCVTACP 300
 QY 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFTGIT 360
 DB 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFTGIT 360
 QY 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFTGIT 360
 DB 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFTGIT 360
 QY 361 ELFPACCKTIFGSLAFPESPDGDPSNTAPLQBPOLQVFETLEBITGYIISAWPDSL 420
 DB 361 IGFACCKTIFGSLAFPESPDGDPSNTAPLQBPOLQVFETLEBITGYIISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLGRLSRLGSGIALIHHNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLGRLSRLGSGIALIHHNTHLCFVHTV 480
 QY 481 PMDQLFRNPHQALHTANRPEDECEVGEGLACHQLCARGCHGWPGPTQCVNSQFLRQEC 540
 DB 481 PMDQLFRNPHQALHTANRPEDECEVGEGLACHQLCARGCHGWPGPTQCVNSQFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAARCLPCHPECCPQNSVTCFGEADQCVACHAYDPPCVARC 600
 DB 541 VEECRVLOGLPREYVNAARCLPCHPECCPQNSVTCFGEADQCVACHAYDPPCVARC 600
 QY 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVVDLDKGPAPORASPLTSISAVVG 660
 DB 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVVDLDKGPAPORASPLTSISAVVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAQMRILKEQYI 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAQMRILKEQYI 720
 QY 721 KNSKRIGITEL-TYKKGWIDGDNVVKIPVAKYLRNTPSPKAKETLIDAVYMAAGVGS 779
 DB 721 R-KVKVLGSGAGFTYKGIWIDGDNVVKIPVAKYLRNTPSPKAKETLIDAVYMAAGVGS 779

QY 780 PYVSRLLGICLTSTVQVLTQMPYGCGLDHRNRCGLSCDILLNMQOIAKMSYLEDV 839
Db 780 PYVSRLLGICLTSTVQVLTQMPYGCGLDHRNRCGLSCDILLNMQOIAKMSYLEDV 839
QY 840 RLVRDLAARNVLYKSPNHYKIDFGIARLLDDETEYADGKVPKIMMALESTLRERF 899
Db 840 RLVRDLAARNVLYKSPNHYKIDFGIARLLDDETEYADGKVPKIMMALESTLRERF 899
QY 900 THSDVMSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVXCM 959
Db 900 THSDVMSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVXCM 959
QY 960 MIDSECRPRELIVSEFSRMAPRQRFVVIQWEDIGPASPILSTYRSLDEDDMDGLVD 1019
Db 960 MIDSECRPRELIVSEFSRMAPRQRFVVIQWEDIGPASPILSTYRSLDEDDMDGLVD 1019
QY 1020 AEEYLVPOQGFPCPDPAFGAGVYHRRHRSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1079
Db 1020 AEEYLVPOQGFPCPDPAFGAGVYHRRHRSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1079
QY 1080 GAGSDVFDGDLGMAKAGIQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
Db 1080 GAGSDVFDGDLGMAKAGIQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
QY 1140 VNQPDVPRQPPSPREGPLPAAPAPGATLERAKTLSPGKNGVYKVFARFAGAVENPEYLT 1199
Db 1140 VNQPDVPRQPPSPREGPLPAAPAPGATLERAKTLSPGKNGVYKVFARFAGAVENPEYLT 1199
QY 1200 QGGAPQPHPPPAFSPAFDNLVYWDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1200 QGGAPQPHPPPAFSPAFDNLVYWDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 2
AAE12130 ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; Apc;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Region Location/Qualifiers
FT 774..782
FT /note="Antigenic epitope"
XX
PN MO200168677-A2.
XX
PD 20-SEP-2001.
XX
PF 16-MAR-2001; 2001MO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2001-616284/71.
XX
DR N-PSDB; AAD19731.
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX
PS Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterized by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6646; DB 22; Length 1255;
Best Local Similarity 97.7%; Pred. No. 0; Mismatches 20; Indels 2; Gaps 2;
Matches 1227; Conservative 7;
QY 1 MEIAALCRMGILLALLPQGAATQVCTGDMKRLRPAPETHLDMRLHLYGQCVVQNTL 60
Db 1 MEIAALCRMGILLALLPQGAATQVCTGDMKRLRPAPETHLDMRLHLYGQCVVQNTL 60
QY 61 ELTYLPTNASLSEFLDIOIEVQSVYLIANQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSEFLDIOIEVQSVYLIANQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIOBPOLCYODTILMKOIFHKNOOLA 180
Db 121 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIOBPOLCYODTILMKOIFHKNOOLA 180
QY 122 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIOBPOLCYODTILMKOIFHKNOOLA 180
Db 122 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIOBPOLCYODTILMKOIFHKNOOLA 180
QY 181 LTLIDTNRBRACHPCSPMCKSGRCWSESEDCQSLTRTVACGACARCGPLPTDCHEQC 240
Db 181 LTLIDTNRBRACHPCSPMCKSGRCWSESEDCQSLTRTVACGACARCGPLPTDCHEQC 240
QY 241 AAGCTGPRXSDCLALHNHSGICEHLPAVYNTDFFSEMPREGRYTTGASCYVACP 300
Db 241 AAGCTGPRXSDCLALHNHSGICEHLPAVYNTDFFSEMPREGRYTTGASCYVACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARCYGLMQYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARCYGLMQYIKANSKFIGIT 360
QY 361 ELIEFAGCKKIFGSLAFLEPSFDGPASTATLQEPOLQVFETLEITGYLYISAMPDLP 420
Db 361 ELIEFAGCKKIFGSLAFLEPSFDGPASTATLQEPOLQVFETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNTQVIRGRILHNGAYSLTQIGISWGLRSLRELSGLALIHNTNLCFHTV 480
Db 421 DLSVFQNTQVIRGRILHNGAYSLTQIGISWGLRSLRELSGLALIHNTNLCFHTV 480
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGALCHQICARHGMGPPTCVVCSQGLRQEC 540
Db 481 PWDQLFRNPQALLHTANRPEDECVGEGALCHQICARHGMGPPTCVVCSQGLRQEC 540
QY 541 VEEGRVLOGLPREYVYARHCPCPECOPOQNGSTTCGPEADQCVACAHKDPFCYARC 600
Db 541 VEEGRVLOGLPREYVYARHCPCPECOPOQNGSTTCGPEADQCVACAHKDPFCYARC 600
QY 601 PSQVYKPDLSYPIKPFDEBAGQPCPINCTHSCVDDLDGCAEGRASLTSIVSAVVG 660
Db 601 PSQVYKPDLSYPIKPFDEBAGQPCPINCTHSCVDDLDGCAEGRASLTSIVSAVVG 660
QY 661 ILTVVVGAVGILIKRQOKIRKTYTRRLLOETLEVEPLTPSGAMPNOQMRILKEQYI 720
Db 661 ILTVVVGAVGILIKRQOKIRKTYTRRLLOETLEVEPLTPSGAMPNOQMRILKEQYI 720

```

QY 721 KANSKFTGITEI..TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAVGVS 779
DB 721 R-KVKYLVGSGAFCTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAVGVS 779
QY 780 PYVSRLLIGICLTSTVQVLTQIMPYGCLLHVRENRRGLSQDILLNWCQIJAQMSYLEDV 839
DB 780 PYVSRLLIGICLTSTVQVLTQIMPYGCLLHVRENRRGLSQDILLNWCQIJAQMSYLEDV 839
QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETEHYADGSKYPIKMALESILRRRF 899
DB 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETEHYADGSKYPIKMALESILRRRF 899
QY 900 THQSDVWSYGVATWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVWCM 959
DB 900 THQSDVWSYGVATWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVWCM 959
QY 960 MIDSECRPFRELYSEFSRWARDPQRFVVIQNEDELGPASPLDSTFYRSILBEDDMDGLVD 1019
DB 960 MIDSECRPFRELYSEFSRWARDPQRFVVIQNEDELGPASPLDSTFYRSILBEDDMDGLVD 1019
QY 1020 AEEYLVPOQGFPCDDPAPGAGMWHHRHSSSTRSGGDLTLGLEPSEBEARSPPLAPSE 1079
DB 1020 AEEYLVPOQGFPCDDPAPGAGMWHHRHSSSTRSGGDLTLGLEPSEBEARSPPLAPSE 1079
QY 1080 GAGSDVFDGDLGMAKAGLSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
DB 1080 GAGSDVFDGDLGMAKAGLSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
QY 1140 VNQPDVTPQPPSPREGPLPAPAPAGATLERAKTLSFGKNGYVKDVAFAGAVENBEYLTP 1199
DB 1140 VNQPDVTPQPPSPREGPLPAPAPAGATLERAKTLSFGKNGYVKDVAFAGAVENBEYLTP 1199
QY 1200 QCGAAPQHPHPAPSPAFDNLVYWDQDPFERGAPSTFKGTPTAENPEYLGLDVEY 1255
DB 1200 QCGAAPQHPHPAPSPAFDNLVYWDQDPFERGAPSTFKGTPTAENPEYLGLDVEY 1255

```

RESULT 3
AAB60167 standard; Protein; 1255 AA.

```

XX AAB60167;
AC AAB60167;
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
OS Homo sapiens.
OS Synthetic.
PN WO200100244-A2.
PD 04-JAN-2001.
PF 23-JUN-2000; 2000WO-US17229.
XX 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX (GETH ) GENENTECH INC.
PA Erickson S, Schwall R;
PI WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

```

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA:

Query Match 97.5%; Score 6646; DB 22; Length 1255;

Best Local Similarity 97.7%; Pred. No. 0;
Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2

```

QY 1 MELAALCRNGILLALLPFGAASVQVCTGDMRLRPASETHLMDLRHLVQGCQVVGNTL 60
DB 1 MELAALCRNGILLALLPFGAASVQVCTGDMRLRPASETHLMDLRHLVQGCQVVGNTL 60
QY 61 ELTYLPTNASLSEFLQDIQEVQGVYLIANQVQVPLQRLIVRGTLPEFNVALAVLNG 120
DB 61 ELTYLPTNASLSEFLQDIQEVQGVYLIANQVQVPLQRLIVRGTLPEFNVALAVLNG 120
QY 121 DPLNNTPTVTGASPGGLARELQLRSLEILKGVLLQRPQLCYQPTILMKDIFHKQNOLA 180
DB 121 DPLNNTPTVTGASPGGLARELQLRSLEILKGVLLQRPQLCYQPTILMKDIFHKQNOLA 180
QY 181 LTLIDNTRSRACPCSPMKGSRGWSSESDCSLTRYCAGGACARCKPLPTDCCHEQC 240
DB 181 LTLIDNTRSRACPCSPMKGSRGWSSESDCSLTRYCAGGACARCKPLPTDCCHEQC 240
QY 241 AAGCTGKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNBRGYTFASCTYACP 300
DB 241 AAGCTGKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNBRGYTFASCTYACP 300
QY 244 AAGCTGKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNBRGYTFASCTYACP 300
DB 244 AAGCTGKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNBRGYTFASCTYACP 300
QY 301 YNYLSTVSGCTLVCPHNOEYVAEDGTCRCEKSPCARVCGYGMQVIXANSKPIGTT 360
DB 301 YNYLSTVSGCTLVCPHNOEYVAEDGTCRCEKSPCARVCGYGMQVIXANSKPIGTT 360
QY 361 ELERAGCKKIFGSLAPFESFDDPASNTAPLOPELOVFEETLGYLYISAWPDSLIP 420
DB 361 ELERAGCKKIFGSLAPFESFDDPASNTAPLOPELOVFEETLGYLYISAWPDSLIP 420
QY 361 IOBRAGCKKIFGSLAPFESFDDPASNTAPLOPELOVFEETLGYLYISAWPDSLIP 420
DB 361 IOBRAGCKKIFGSLAPFESFDDPASNTAPLOPELOVFEETLGYLYISAWPDSLIP 420
QY 421 DLSYFQNLQVIRGRILHNAGYSLTLOGLSISWGLSLSELGSGALLHHNHLCTVHTV 480
DB 421 DLSYFQNLQVIRGRILHNAGYSLTLOGLSISWGLSLSELGSGALLHHNHLCTVHTV 480
QY 481 PMDOLFENPHQALLHTANRPEDECEVGEGLACHQLCARGHCMGPGPTQCVNCSQFLRGQEC 540
DB 481 PMDOLFENPHQALLHTANRPEDECEVGEGLACHQLCARGHCMGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLIQGLPREYVNAHCLPCHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
DB 541 VEECRVLIQGLPREYVNAHCLPCHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
QY 541 VEECRVLIQGLPREYVNAHCLPCHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
DB 541 VEECRVLIQGLPREYVNAHCLPCHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
QY 601 PSQVPLDSTFYRSILBEDDMDGLVD 1019
DB 601 PSQVPLDSTFYRSILBEDDMDGLVD 1019
QY 661 ILVYVVGAVGFIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKEQYI 720
DB 661 ILVYVVGAVGFIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKEQYI 720
QY 721 KANSKFTGITEI..TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAVGVS 779
DB 721 KANSKFTGITEI..TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAVGVS 779
QY 780 PYVSRLLIGICLTSTVQVLTQIMPYGCLLHVRENRRGLSQDILLNWCQIJAQMSYLEDV 839
DB 780 PYVSRLLIGICLTSTVQVLTQIMPYGCLLHVRENRRGLSQDILLNWCQIJAQMSYLEDV 839
QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETEHYADGSKYPIKMALESILRRRF 899
DB 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETEHYADGSKYPIKMALESILRRRF 899

```


Db 840 RLVRHDLAARNLVKSPNKHITDFGLARLLIDIDETEVHADGCKVPIKMALESILRRR 899
 QY THSDPWSYGVTVNLMFPGAKPYDGI PAREIPDLLEKGERLPDPICITIDVYMIWKCW 959
 Db THSDPWSYGVTVNLMFPGAKPYDGI PAREIPDLLEKGERLPDPICITIDVYMIWKCW 959
 QY 960 MIDSECRPRFRELVSFGRMARDPQRFVITONEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
 Db 960 MIDSECRPRFRELVSFGRMARDPQRFVITONEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
 QY 1020 ABEYLVPPQGFCCPDPAAGAGMWHHRSSSTSSGGGDLTLGLPESEERAPSPPLAPSE 1079
 Db 1020 ABEYLVPPQGFCCPDPAAGAGMWHHRSSSTSSGGGDLTLGLPESEERAPSPPLAPSE 1079
 QY 1080 GAGSDVFPDGLGMAKGLQSLPTHDPSPLOKYSDDPTVLPSETDGYVAPLTCSPQPEX 1139
 Db 1080 GAGSDVFPDGLGMAKGLQSLPTHDPSPLOKYSDDPTVLPSETDGYVAPLTCSPQPEX 1139
 QY 1140 VNQDPVRPQPPSPREGPLPARPAGATLERAKTILSPKNGVVDVPAFGAVENPEYLP 1199
 Db 1140 VNQDPVRPQPPSPREGPLPARPAGATLERAKTILSPKNGVVDVPAFGAVENPEYLP 1199
 QY 1200 QGGAAPQPHPPAPSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1200 QGGAAPQPHPPAPSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX 23-APR-2002 (first entry)
 DT Human HER2 (ErbB2) polypeptide.
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glioma disorder; astrocytoma disorder; hypothalamic disorder;
 KW glandular disorder; macropneumal disorder; epithelial disorder;
 KW stromal disorder; blastocoele disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOMSKI M.
 XX
 PI Erickson S, Schwall R, Sliwowski M;
 XX WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor. ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytic, hypothalamic, glandular, macropneumal,
 CC epithelial, stromal, blastocoele, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.5%; Score 6646; DB 23; Length 1255;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2;
 QY 1 MELALCRWGLLALLPFGAASITQCTGDMKRLPASPETHDMLRHLYQGCGVQGNL 60
 Db 1 MELALCRWGLLALLPFGAASITQCTGDMKRLPASPETHDMLRHLYQGCGVQGNL 60
 QY 61 ELTYLPTNASTLFIQDIQEVQGYVLIHNOVROYVLPQRLRIVRGTLFEDNVYALVINDG 120
 Db 61 ELTYLPTNASTLFIQDIQEVQGYVLIHNOVROYVLPQRLRIVRGTLFEDNVYALVINDG 120
 QY 121 DPLNNTTPVTCASGSGRLREQLRLSTELILKGVLIQNPOLCYQDTILMKDITHKNNQOLA 180
 Db 121 DPLNNTTPVTCASGSGRLREQLRLSTELILKGVLIQNPOLCYQDTILMKDITHKNNQOLA 180
 QY 181 LTLIDITNRSRACHPSCSMCKSGRSMGSSSEDCQSLRTVTCAGGACARCKGPLPTDCHEQC 240
 Db 181 LTLIDITNRSRACHPSCSMCKSGRSMGSSSEDCQSLRTVTCAGGACARCKGPLPTDCHEQC 240
 QY 241 AAGCTGPRHSDCLALCFPNHNSGICELHCPALVYNTDFTFSMNPBGRYTFGASCYTAC 300
 Db 241 AAGCTGPRHSDCLALCFPNHNSGICELHCPALVYNTDFTFSMNPBGRYTFGASCYTAC 300
 QY 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
 QY 361 ELBPAQCKTIFGSLAFIPESFDDGPASNTAPLQPEQVETLEETITGYLYISAMPDSIP 420
 Db 361 IQEPAGCKTIFGSLAFIPESFDDGPASNTAPLQPEQVETLEETITGYLYISAMPDSIP 420
 QY 421 DLSVFONLQVTRGRILHNGAYSLTLQGLSMILGSLRELSSGLALIHNTHTLCFVHTV 480
 Db 421 DLSVFONLQVTRGRILHNGAYSLTLQGLSMILGSLRELSSGLALIHNTHTLCFVHTV 480
 QY 481 FMDQLFRNPQALHTANRPEDCEVSGELACHQJLCARGHCGPQTQCVNCSQFLRGQEC 540
 Db 481 FMDQLFRNPQALHTANRPEDCEVSGELACHQJLCARGHCGPQTQCVNCSQFLRGQEC 540
 QY 541 VEECRVYQGLPREVYNAHCLPHPEQOPNGSVTCFGBPADQCVACAHKXDPFCVABC 600
 Db 541 VEECRVYQGLPREVYNAHCLPHPEQOPNGSVTCFGBPADQCVACAHKXDPFCVABC 600
 QY 601 PSQVAPDLSYMPIMKPFDEGACQPCINCTHSCVDLDDKGCABEGRASPLTIVAVVG 660
 Db 601 PSQVAPDLSYMPIMKPFDEGACQPCINCTHSCVDLDDKGCABEGRASPLTIVAVVG 660
 QY 661 ILVVVVGAVFGILIKRQOKIRKTYRRLQETELVEPLTPSGAMPNOQWILKETEL 720
 Db 661 ILVVVVGAVFGILIKRQOKIRKTYRRLQETELVEPLTPSGAMPNOQWILKETEL 720
 QY 721 KANSKFIGITEL-TVYGIWIPGENVKIPVAIKVRENTSPKANEILDEAVYVAGVS 779
 Db 721 R-KYKVLGSGAFGIVKGIWIPGENVKIPVAIKVRENTSPKANEILDEAVYVAGVS 779
 QY 780 PYVSRLLIGICLTSTVQLVLTQJLMPYGCILDHVRENRLGSDILNMCQIAGKMSYLEDV 839

```

Db      780 PYVSRLLGICLTSTVQLVLTQLMFPGCLLDHVRNRRGLSGODLLMCMQIAKMSYLEEV 839
Qy      840 FLVHRDLAARNVLYKSPNHVYKTTDFGLARLLDDETEYHAGDKVPIKMALESILRRRF 899
Db      840 FLVHRDLAARNVLYKSPNHVYKTTDFGLARLLDDETEYHAGDKVPIKMALESILRRRF 899
Qy      900 THOSDWSYGVTVLMTFGAKPYDGIIPABEIPDLLEKGRLLPQPPICITDVMIMVKCM 959
Db      900 THOSDWSYGVTVLMTFGAKPYDGIIPABEIPDLLEKGRLLPQPPICITDVMIMVKCM 959
Qy      960 MIDSECRFRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
Db      960 MIDSECRFRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
Qy      1020 AEEYLVPOQGFCCDPAPAGAGWVHHRRSSSTRSGGDLTLLESEEBAPSPPLAPSE 1079
Db      1020 AEEYLVPOQGFCCDPAPAGAGWVHHRRSSSTRSGGDLTLLESEEBAPSPPLAPSE 1079
Qy      1080 GAGSDVPDGDLMGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQPEY 1139
Db      1080 GAGSDVPDGDLMGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQPEY 1139
Qy      1140 VNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVADVAPFGAVENPEYLTP 1199
Db      1140 VNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVADVAPFGAVENPEYLTP 1199
Qy      1200 QGGAAPQHPPPAPSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVPV 1255
Db      1200 QGGAAPQHPPPAPSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 5
AA001111 ID AA001111 standard; Protein; 1255 AA.
AC      AA001111;
DT      01-JAN-1997 (first entry)
DE      HER-2/neu protein.
XX      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW      breast cancer; ovary cancer; colon cancer; lung cancer;
KM      prostate cancer; immunisation; tumour; vaccine; vector.
XX      Homo sapiens.
OS
FH      Key Location/Qualifiers
FT      676..1255
FT      /label= "intracellular domain"
FT      /note= "claimed domain, useful for immunisation"
XX      MO9630514-A1.
XX      PD 03-OCT-1996.
XX      PE 28-MAR-1996; 96MO-US01689.
XX      PR 31-MAR-1995; 95US-0414417.
XX      PA (UNIW ) UNIV WASHINGTON.
XX      PI Cheever MA, Disis ML;
XX      DR N-PSDB; AAT40739.
XX      WIPI: 1996-455361/45.
XX      PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX      treatment of malignancies with which the HER-2/neu oncogene is
XX      associated
PS      Claim 2; 'page 56-61, 71pp; English.

```

```

XX      Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also AAT40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transfected host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX
SQ      Sequence 1255 AA:
Query Match 97.4%; Score 6640; DB 17; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;
Qy      1 MELAALCRWGLLALLPFGAASVQVCTGTDMLRLPASETHLDMRLHYQGCQVVGNTL 60
Db      1 MELAALCRWGLLALLPFGAASVQVCTGTDMLRLPASETHLDMRLHYQGCQVVGNTL 60
Qy      61 ELTYLPNASTLFQDIQEVQGVLLAHNOVRVPLQRLIRVGTQLFEDNYALAVLUNG 120
Db      61 ELTYLPNASTLFQDIQEVQGVLLAHNOVRVPLQRLIRVGTQLFEDNYALAVLUNG 120
Qy      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLLIQRNQOLCYQDTILWKDIFHKNNQLA 180
Db      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLLIQRNQOLCYQDTILWKDIFHKNNQLA 180
Qy      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLLIQRNQOLCYQDTILWKDIFHKNNQLA 180
Db      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLLIQRNQOLCYQDTILWKDIFHKNNQLA 180
Qy      181 LTLIDNRSRACHPCSCMCKSGSCWSESSEDCQSLRTYVACAGGACARCPPLDCCHEQC 240
Db      181 LTLIDNRSRACHPCSCMCKSGSCWSESSEDCQSLRTYVACAGGACARCPPLDCCHEQC 240
Qy      181 LTLIDNRSRACHPCSCMCKSGSCWSESSEDCQSLRTYVACAGGACARCPPLDCCHEQC 240
Db      181 LTLIDNRSRACHPCSCMCKSGSCWSESSEDCQSLRTYVACAGGACARCPPLDCCHEQC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTPFGASCYACAP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTPFGASCYACAP 300
Qy      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTPFGASCYACAP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTPFGASCYACAP 300
Qy      301 YNYLSTDVSGCTLVCPAHNGEVTAEADGTORCEKSKPCARVCYGLQMYIKANSKFIGIT 360
Db      301 YNYLSTDVSGCTLVCPAHNGEVTAEADGTORCEKSKPCARVCYGLQMYIKANSKFIGIT 360
Qy      301 YNYLSTDVSGCTLVCPAHNGEVTAEADGTORCEKSKPCARVCYGLQMYIKANSKFIGIT 360
Db      301 YNYLSTDVSGCTLVCPAHNGEVTAEADGTORCEKSKPCARVCYGLQMYIKANSKFIGIT 360
Qy      361 ELFPAGCKKIFGSLAFPSFDDSPASNTAPLOPQVFEETLEETLGYLYISAMPDLSIP 420
Db      361 ELFPAGCKKIFGSLAFPSFDDSPASNTAPLOPQVFEETLEETLGYLYISAMPDLSIP 420
Qy      361 IQEPAGCKKIFGSLAFPSFDDSPASNTAPLOPQVFEETLEETLGYLYISAMPDLSIP 420
Db      361 IQEPAGCKKIFGSLAFPSFDDSPASNTAPLOPQVFEETLEETLGYLYISAMPDLSIP 420
Qy      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLGCVHTV 480
Db      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLGCVHTV 480
Qy      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLGCVHTV 480
Db      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLGCVHTV 480
Qy      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Db      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Qy      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Db      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Qy      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Db      481 PMDOLFNNPHQALLHTRNPEDECVGEGLACHQLCARGHCMKGRPTQCVNGSQFLRGQEC 540
Qy      541 VEEGRVYQGLPREVYVARHCLPCHPCQOPONGSVTCFGEADOCVACAHYKDPFCVVARC 600
Db      541 VEEGRVYQGLPREVYVARHCLPCHPCQOPONGSVTCFGEADOCVACAHYKDPFCVVARC 600
Qy      541 VEEGRVYQGLPREVYVARHCLPCHPCQOPONGSVTCFGEADOCVACAHYKDPFCVVARC 600
Db      541 VEEGRVYQGLPREVYVARHCLPCHPCQOPONGSVTCFGEADOCVACAHYKDPFCVVARC 600
Qy      601 PSQVYKPLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPQASPLTISIAVVG 660
Db      601 PSQVYKPLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPQASPLTISIAVVG 660
Qy      601 PSQVYKPLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPQASPLTISIAVVG 660
Db      601 PSQVYKPLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPQASPLTISIAVVG 660
Qy      661 ILVYVYGVVFGILIRROOKIRKYMRRLLQSTELVEPLTSGAMPNOAQRILKEQVI 720
Db      661 ILVYVYGVVFGILIRROOKIRKYMRRLLQSTELVEPLTSGAMPNOAQRILKEQVI 720
Qy      661 ILVYVYGVVFGILIRROOKIRKYMRRLLQSTELVEPLTSGAMPNOAQRILKEQVI 720
Db      661 ILVYVYGVVFGILIRROOKIRKYMRRLLQSTELVEPLTSGAMPNOAQRILKEQVI 720
Qy      721 KANSKFIGITEL-TYVKGIMIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Db      721 KANSKFIGITEL-TYVKGIMIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Qy      721 KANSKFIGITEL-TYVKGIMIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Db      721 KANSKFIGITEL-TYVKGIMIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Qy      760 PYVSRLLGICLTSTVQLVLTQLMFPGCLLDHVRNRRGLSGODLLMCMQIAKMSYLEEV 839
Db      760 PYVSRLLGICLTSTVQLVLTQLMFPGCLLDHVRNRRGLSGODLLMCMQIAKMSYLEEV 839
Qy      780 PYVSRLLGICLTSTVQLVLTQLMFPGCLLDHVRNRRGLSGODLLMCMQIAKMSYLEEV 839
Db      780 PYVSRLLGICLTSTVQLVLTQLMFPGCLLDHVRNRRGLSGODLLMCMQIAKMSYLEEV 839

```

QY 840 RLVHRDLAARNVLYKSPNNHYKITDFFGLARLLDIDETFEYHADGKVPITKMALESILRRRF 899
 DB 840 RLVHRDLAARNVLYKSPNNHYKITDFFGLARLLDIDETFEYHADGKVPITKMALESILRRRF 899
 QY 900 THOSDWSYGVTVWELMTFGAKFYDGIIPAREIPDLEKEKERLPQPICTIDVYIMWKW 959
 DB 900 THOSDWSYGVTVWELMTFGAKFYDGIIPAREIPDLEKEKERLPQPICTIDVYIMWKW 959
 QY 960 MIDSECRPRFRELVSERSRARDPORFVVIIONEDLGPASPLDSTYRSLDDDMGLVD 1019
 DB 960 MIDSECRPRFRELVSERSRARDPORFVVIIONEDLGPASPLDSTYRSLDDDMGLVD 1019
 QY 1020 AEELVLPQGFPCPDPAAGAGVWHRHRSSTRSGGGLTLGLEPSEEARPSPLAPSE 1079
 DB 1020 AEELVLPQGFPCPDPAAGAGVWHRHRSSTRSGGGLTLGLEPSEEARPSPLAPSE 1079
 QY 1080 GAGSDVFDGLMGAAKGLQSLFTHDPSLQRYSEDPTVPLPSETDGYVAALTCSPQPEY 1139
 DB 1080 GAGSDVFDGLMGAAKGLQSLFTHDPSLQRYSEDPTVPLPSETDGYVAALTCSPQPEY 1139
 QY 1140 VNCPDVRPQPPSPREGPLPARAPAGATLEAKTLSPGKGVVQVFAFGAVENPEYLTLP 1199
 DB 1140 VNCPDVRPQPPSPREGPLPARAPAGATLEAKTLSPGKGVVQVFAFGAVENPEYLTLP 1199
 QY 1200 QGGAAPQPPHPPAPSPAFDNLVYWDODPPERGAPOSTFEKGTPTAENPEYLGIDYVPV 1255
 DB 1200 QGGAAPQPPHPPAPSPAFDNLVYWDODPPERGAPOSTFEKGTPTAENPEYLGIDYVPV 1255

RESULT 6

AAW92406 standard; Protein: 1255 AA.

AAW92406;

21-APR-1999 (first entry)

Human HER-2/neu oncogene protein.

HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;

malignancy; treatment; tumour.

Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers
 676..1255
 /note="region which elicits immune response"

US5869445-A.

09-FEB-1999.

01-APR-1996;

01-APR-1996;

01-APR-1996;

17-MAR-1993;

12-AUG-1993;

31-MAR-1995;

(UNIM) UNIV WASHINGTON.

Cheever MA, Disis ML;

WPI, 1999-152835/13.

N-PSDB; AAX01912.

Use of HER-2/neu polypeptides - for eliciting an immune response to

an HER-2/neu associated malignancy, particularly for treating or

preventing tumours

Claim 3; Column 31-38; 26pp; English.

This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 20; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MEALALCWGLLLLLPPGAASVTCGTGTMKRLPSPETHLDMRLHYOGGVQVQNL 60
 DB 1 MEALALCWGLLLLLPPGAASVTCGTGTMKRLPSPETHLDMRLHYOGGVQVQNL 60
 QY 61 ELTYLPNASLSFLQDIQEVQGYLLAHNOVRQVPLQRLRYRGTLFEDNALAVLDNG 120
 DB 61 ELTYLPNASLSFLQDIQEVQGYLLAHNOVRQVPLQRLRYRGTLFEDNALAVLDNG 120
 QY 121 DPLNNTTPTGASPGGLRELQRLSLTEILKGYLLIQNPQLCYQDTILMKDIFHKXNOLA 180
 DB 121 DPLNNTTPTGASPGGLRELQRLSLTEILKGYLLIQNPQLCYQDTILMKDIFHKXNOLA 180
 QY 181 LTLIDTNRSPACHPCSPCKSGRCWGSSEDCQSLRTVACAGGACCKGPLPTDCHEQC 240
 DB 181 LTLIDTNRSPACHPCSPCKSGRCWGSSEDCQSLRTVACAGGACCKGPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLALPHNHSIGCEALCPALVYNTDTPESMNPREGRTFFGASCTACP 300
 DB 241 AAGCTGPKHSDCLALPHNHSIGCEALCPALVYNTDTPESMNPREGRTFFGASCTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNOEVTAEQDTCRCKSPKACVCGYGLMOYIKANSKEIGIT 360
 DB 301 YNYLSTDVGSCTLVCPHLNOEVTAEQDTCRCKSPKACVCGYGLMOYIKANSKEIGIT 360
 QY 361 ELFPAGCKTIFGSLAFIPESFDGPAANTPLQPELOQVETLEETLGYLXISAMPDISLP 420
 DB 361 ELFPAGCKTIFGSLAFIPESFDGPAANTPLQPELOQVETLEETLGYLXISAMPDISLP 420
 QY 421 DLSVFNQNLQVIRGRILHNQAVSLTLQGLISWLGRLSRLSGSLALIHNNHLCFVHTV 480
 DB 421 DLSVFNQNLQVIRGRILHNQAVSLTLQGLISWLGRLSRLSGSLALIHNNHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPBDECVGSLACHQLCARHGWGPGPQCVNCSQFLRGQEC 540
 DB 481 PWDQLFRNPHQALLHTANRPBDECVGSLACHQLCARHGWGPGPQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPEADQCVACAHYKDPFCVABC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPEADQCVACAHYKDPFCVABC 600
 QY 601 PSGVCPDLSTYMPFKFPDEBGAQPCPCINCTHSCVDLDKCAEGRASPSTYSAYVG 660
 DB 601 PSGVCPDLSTYMPFKFPDEBGAQPCPCINCTHSCVDLDKCAEGRASPSTYSAYVG 660
 QY 661 ILVYVGLGVFGIILIKRQCKIRKYTRKLQETELVEPLTPSGAMPNOAKRIKEGYI 720
 DB 661 ILVYVGLGVFGIILIKRQCKIRKYTRKLQETELVEPLTPSGAMPNOAKRIKEGYI 720
 QY 721 KANSKFIGITEL-TVYKGIWIPDGENVYKIVAILKURENTSPANKELIDEAYVMAVGS 779
 DB 721 KANSKFIGITEL-TVYKGIWIPDGENVYKIVAILKURENTSPANKELIDEAYVMAVGS 779
 QY 780 PYVSRLLGICLTSIVQLVQLMPGCLDHEVRENRRLSODLLANCMQIAGMSLEDEY 839
 DB 780 PYVSRLLGICLTSIVQLVQLMPGCLDHEVRENRRLSODLLANCMQIAGMSLEDEY 839
 QY 840 RLVHRDLAARNVLYKSPNNHYKITDFFGLARLLDIDETFEYHADGKVPITKMALESILRRRF 899
 DB 840 RLVHRDLAARNVLYKSPNNHYKITDFFGLARLLDIDETFEYHADGKVPITKMALESILRRRF 899

QY 900 THSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGGERLPOPICTIDVYMIWKCM 959
 Db 900 THSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGGERLPOPICTIDVYMIWKCM 959
 QY 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMDGLVD 1019
 Db 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMDGLVD 1019
 QY 1020 AEEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEBAPRSLAPSE 1079
 Db 1020 AEEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEBAPRSLAPSE 1079
 QY 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 Db 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1140 VNOPDVRPOPSPREPLPAAPAGATLERAKTSLSPGKGVYKQVFAFGAVENPEYLTP 1139
 Db 1140 VNOPDVRPOPSPREPLPAAPAGATLERAKTSLSPGKGVYKQVFAFGAVENPEYLTP 1139
 QY 1200 QGGAPQPHPPAPSPAFDNLVYWDPPERGAPSPSTFKGPTAENPEYLGLDVPV 1255
 Db 1200 QGGAPQPHPPAPSPAFDNLVYWDPPERGAPSPSTFKGPTAENPEYLGLDVPV 1255

RESULT 7
 AAB21198
 ID AAB21198 standard; protein; 1255 AA.
 AC AAB21198;
 XX
 XX 12-JAN-2001 (first entry)
 DT
 XX
 DE Human HER-2/neu protein.
 XX
 XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KM breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KM colon cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200044899-A1.
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000MO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 PA (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI: 2000-505976/45.
 DR N-PSDB; AAA89736.
 XX
 PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 52; Fig 7; 128pp; English.
 XX
 CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.4%; Score 6640; DB 21; Length 1255;
 Best Local Similarity 97.5%; Pred No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2.

QY 1 MELALCRMGLLALLPGAASTQVCTGDMKRLPASPEHLDMLRLYQGCQVQGNL 60
 Db 1 MELALCRMGLLALLPGAASTQVCTGDMKRLPASPEHLDMLRLYQGCQVQGNL 60
 QY 61 ELTYLPTNASLFLDIOIEVQGVYLIANOVROYVLOGLRTVRGQLPEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLDIOIEVQGVYLIANOVROYVLOGLRTVRGQLPEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELOQLSLTEILKGGVLIQRPOLCYODTILMKDIFKNNOLA 180
 Db 121 DPLNNTPTVTGASPGGLRELOQLSLTEILKGGVLIQRPOLCYODTILMKDIFKNNOLA 180
 QY 181 LTLIDNRSRACHPSPSPKSGKSCWSESSDCCSLTIVCAGGACARCKGLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPSPSPKSGKSCWSESSDCCSLTIVCAGGACARCKGLPTDCCHEQC 240
 QY 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVYNTDTPFESMPNREGRYTFGASCVTACP 300
 Db 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVYNTDTPFESMPNREGRYTFGASCVTACP 300
 QY 301 YNYLSTDVQSCTLVCPHLNQEVTAEADGTQRCCKSKPCARVYGLGNQYI KANSKFTGIT 360
 Db 301 YNYLSTDVQSCTLVCPHLNQEVTAEADGTQRCCKSKPCARVYGLGNQYI KANSKFTGIT 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDGPASTAPLOEQLOVEFTLEITGYLYISMPISLP 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGPASTAPLOEQLOVEFTLEITGYLYISMPISLP 420
 QY 421 DLSVFQNLQVIRGRILNHGAYSLTLQGLISWLSGLRSLRELSGLALIHNTHLCEFTV 480
 Db 421 DLSVFQNLQVIRGRILNHGAYSLTLQGLISWLSGLRSLRELSGLALIHNTHLCEFTV 480
 QY 481 PWDQFRNPHQALLHTANRPEDECVGSLACHQICAGHGMGSPPTCVNCSOFLRQEC 540
 Db 481 PWDQFRNPHQALLHTANRPEDECVGSLACHQICAGHGMGSPPTCVNCSOFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPQNGSVTCGFPADQCVACAHKDDPFVYAR 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECOPQNGSVTCGFPADQCVACAHKDDPFVYAR 600
 QY 601 PSQVPRDLSYMPIMKFPDEBGAQPCPINCCHSCVDLDDKGCAPAEORASPLTISVAWG 660
 Db 601 PSQVPRDLSYMPIMKFPDEBGAQPCPINCCHSCVDLDDKGCAPAEORASPLTISVAWG 660
 QY 661 ILLVVLGVVFGIILIKRROOKIRYTRRLLOETELVEPLTPSGAMPNQOMILKEQYI 720
 Db 661 ILLVVLGVVFGIILIKRROOKIRYTRRLLOETELVEPLTPSGAMPNQOMILKEQYI 720
 QY 721 KANSKFTGITEL-TVYKGIWIPDENYKIPVAIKVLRENTSPPANKEILDEAVYMAVGVS 779
 Db 721 KANSKFTGITEL-TVYKGIWIPDENYKIPVAIKVLRENTSPPANKEILDEAVYMAVGVS 779
 QY 780 PYVSRLLGICLSTVQCVLTQMLPQGLLDHVRNRRGLSGQDLINMCKQIAKMSYLEBV 839
 Db 780 PYVSRLLGICLSTVQCVLTQMLPQGLLDHVRNRRGLSGQDLINMCKQIAKMSYLEBV 839
 QY 840 RLVRDIAARNVILVKSBNHYKITDFGLARLDDIDETEHADGSKVPIKMALESILRRRF 899
 Db 840 RLVRDIAARNVILVKSBNHYKITDFGLARLDDIDETEHADGSKVPIKMALESILRRRF 899
 QY 900 THSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGGERLPOPICTIDVYMIWKCM 959
 Db 900 THSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGGERLPOPICTIDVYMIWKCM 959
 QY 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMDGLVD 1019

```

Db      960 MIDSECRPRFRELVSFSSMARDPQRFVVIQNEIDGPRASPLDSTFFYSLSLEDDMDGLVD 1019
Qy      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSE 1079
Db      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSE 1079
Qy      1080 GAGSDVFDGDLGGAAGKGLSLPTHDPSPLORYSDDPTVPLPSETDGVVAPLTCSPQPEX 1139
Db      1080 GAGSDVFDGDLGGAAGKGLSLPTHDPSPLORYSDDPTVPLPSETDGVVAPLTCSPQPEX 1139
Qy      1140 VNQDVRPQPPSPFRSGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAAVENPEYLT 1199
Db      1140 VNQDVRPQPPSPFRSGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAAVENPEYLT 1199
Qy      1200 QGGAAPQPPHPPAFSPAFDNLVYMDODPBERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db      1200 QGGAAPQPPHPPAFSPAFDNLVYMDODPBERGAPSTFKGTPTAENPEYLGIDVPV 1255

```

RESULT 8

AAV84780

ID AAV84780 standard; Protein; 1255 AA.

AAV84780;

08-AUG-2000 (first entry)

Amino acid sequence of the SPICE erbB-2 receptor protein.

SPICE erbB-2 receptor protein; cell transformation disorder; cancer;

tumor cell proliferation; tissue degeneration; arthropathy;

bone resorption; inflammatory disease; degenerative disorder;

wound healing.

Homo sapiens.

MO200020579-A1.

13-APR-2000.

01-OCT-1999; 99MO-CH00912.

02-OCT-1998; 98US-0165192.

(UYMC-) UNIV MCMASTER.

Muller WJ, Siegel PM.

WPI; 2000-303768/26.

N-PSDB; AAL14812.

Claim 3; Fig 2; 60pp; English.

The present sequence represents a SPICE erbB-2 receptor protein. The protein has an in-frame deletion of 16 amino acids, 2 of which are conserved cysteine residues, compared to the unspliced protein. The erbB-2 polynucleotide is used to construct probes for detecting disorders of cell transformation such as cancer. Antibodies to the protein may be used to detect SPICE erbB-2 in a sample. Agents of (e.g. antisense oligonucleotides) which inhibit the expression of SPICE erbB-2 are useful for reducing tumor cell proliferation and treating cancer. Substances which stimulate erbB-2 are useful in which degeneration of tissue occurs, such as arthropathy, bone resorption, inflammatory diseases, degenerative disorders of the central nervous system and wound healing.

Sequence 1255 AA;

```

Query Match      97.4%; Score 6640; DB 21; Length 1255;
Beet Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

Qy      1 MELALCRMGILLALLPPGAASVQVCTGDMKLRPASDETHLDMRLHYQCQVYQNL 60
Db      1 MELALCRMGILLALLPPGAASVQVCTGDMKLRPASDETHLDMRLHYQCQVYQNL 60
Qy      61 ELTYLPTNASLSFLQDIQEVGVVLAHNOVQVPLQRIYRGTQLFEDNVALVLNDG 120
Db      61 ELTYLPTNASLSFLQDIQEVGVVLAHNOVQVPLQRIYRGTQLFEDNVALVLNDG 120
Qy      61 ELTYLPTNASLSFLQDIQEVGVVLAHNOVQVPLQRIYRGTQLFEDNVALVLNDG 120
Db      61 ELTYLPTNASLSFLQDIQEVGVVLAHNOVQVPLQRIYRGTQLFEDNVALVLNDG 120
Qy      121 DPLNNTTPVTGASPGGLRELQSLTEILKGVLLIQRNQOLCYQDTIIMKDIFFHKNOLA 180
Db      121 DPLNNTTPVTGASPGGLRELQSLTEILKGVLLIQRNQOLCYQDTIIMKDIFFHKNOLA 180
Qy      121 DPLNNTTPVTGASPGGLRELQSLTEILKGVLLIQRNQOLCYQDTIIMKDIFFHKNOLA 180
Db      121 DPLNNTTPVTGASPGGLRELQSLTEILKGVLLIQRNQOLCYQDTIIMKDIFFHKNOLA 180
Qy      181 LTLIDTRSRACPCSCMCKGSCWSESEDDQSLTRYCAGGACRCGPLPTDCCHQC 240
Db      181 LTLIDTRSRACPCSCMCKGSCWSESEDDQSLTRYCAGGACRCGPLPTDCCHQC 240
Qy      181 LTLIDTRSRACPCSCMCKGSCWSESEDDQSLTRYCAGGACRCGPLPTDCCHQC 240
Db      181 LTLIDTRSRACPCSCMCKGSCWSESEDDQSLTRYCAGGACRCGPLPTDCCHQC 240
Qy      241 AACCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
Db      241 AACCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
Qy      301 YNYLSTVGSCTLVCPHLNDEVTAEDGTORCEKSPCARVYCYGMOYIKANSKFTGTT 360
Db      301 YNYLSTVGSCTLVCPHLNDEVTAEDGTORCEKSPCARVYCYGMEHLREYRAVTSAN 360
Qy      361 ELSPAGCKKIFGSLAF.PESFDDPASPANAP.OPBOLQVFEETLEITGLVLSAMPDSL 420
Db      361 ELSPAGCKKIFGSLAF.PESFDDPASPANAP.OPBOLQVFEETLEITGLVLSAMPDSL 420
Qy      361 IQEFGAGCKKIFGSLAF.PESFDDPASPANAP.OPBOLQVFEETLEITGLVLSAMPDSL 420
Db      361 IQEFGAGCKKIFGSLAF.PESFDDPASPANAP.OPBOLQVFEETLEITGLVLSAMPDSL 420
Qy      421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGLALIHHTHLCFVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGLALIHHTHLCFVHTV 480
Qy      481 PMDQLFNPHQALHTANPREDECEVGEGLACQOLCARGCMGPGTQCVNCSQFLRGEC 540
Db      481 PMDQLFNPHQALHTANPREDECEVGEGLACQOLCARGCMGPGTQCVNCSQFLRGEC 540
Qy      541 VEECRVLOGLPREYVNAHCLPCHPECPQONSVCFCGEADQCAVAHYXDPFCVARC 600
Db      541 VEECRVLOGLPREYVNAHCLPCHPECPQONSVCFCGEADQCAVAHYXDPFCVARC 600
Qy      601 PSQVAPDLSYMPIKTFDEEGACQPCPINCTHSCVDLDKGCAPABORASPLTSTSAV 660
Db      601 PSQVAPDLSYMPIKTFDEEGACQPCPINCTHSCVDLDKGCAPABORASPLTSTSAV 660
Qy      661 ILVVLGVVFGILIRROQKIRKTYMRRLQETELVEPLPSGAMPQAQWRILKEQYI 720
Db      661 ILVVLGVVFGILIRROQKIRKTYMRRLQETELVEPLPSGAMPQAQWRILKEQYI 720
Qy      721 KANSKFTIGTEL.TYKKGWIPDGEVVK.PVAKYLKRENTSKAKETILDEAYVMAAG 779
Db      721 KANSKFTIGTEL.TYKKGWIPDGEVVK.PVAKYLKRENTSKAKETILDEAYVMAAG 779
Qy      780 PYVSRLLGICLTSTVOLVTLQMPYGLDVAENRGRLSQDLNMCQIAKAGSYLEDV 839
Db      780 PYVSRLLGICLTSTVOLVTLQMPYGLDVAENRGRLSQDLNMCQIAKAGSYLEDV 839
Qy      840 RLYHRDLAANVLYKSPNHYKTTDFGLALLIDITETVHAOGKVPIMWLESTLRFR 899
Db      840 RLYHRDLAANVLYKSPNHYKTTDFGLALLIDITETVHAOGKVPIMWLESTLRFR 899
Qy      900 THQSDVSYGVTVWELMTGAKYDIPAREIPDLLEKGERLPQPPICITIDVYIMVW 959
Db      900 THQSDVSYGVTVWELMTGAKYDIPAREIPDLLEKGERLPQPPICITIDVYIMVW 959
Qy      960 MIDSECRPRFRELVSFSSMARDPQRFVVIQNEIDGPRASPLDSTFFYSLSLEDDMDGLVD 1019
Db      960 MIDSECRPRFRELVSFSSMARDPQRFVVIQNEIDGPRASPLDSTFFYSLSLEDDMDGLVD 1019
Qy      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSE 1079

```

Db 1020 AEEYLVPQGGFFCCDPAPAGAGVHHRSSSTRSGGGDLTLGLEPSEEARPRPLAPSE 1079
QY 1080 GAGSDVFDGDLGMGAAGKIQSLPTHDSPLQRYSDPTVPLPSETDGVAPLTGSPQPEY 1139
Db 1080 GAGSDVFDGDLGMGAAGKIQSLPTHDSPLQRYSDPTVPLPSETDGVAPLTGSPQPEY 1139
QY 1140 VAGQDVPRPQPSRPGSLPAARPAATLERAKTSLPGKNGVYKDFAFGAVENPEYLT 1199
Db 1140 VAGQDVPRPQPSRPGSLPAARPAATLERAKTSLPGKNGVYKDFAFGAVENPEYLT 1199
QY 1200 QCGAAPQPHPPAFSPAFDNLVYWDQDPPEKAPSTPKGTPTAENPEYIGLDVAV 1255
Db 1200 QCGAAPQPHPPAFSPAFDNLVYWDQDPPEKAPSTPKGTPTAENPEYIGLDVAV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein, 1255 AA.

XX AAB85458;
AC AAB85458;
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
OS Homo sapiens.
PN MO200153463-A2.
XX 26-JUL-2001.
PD 19-JAN-2001; 2001MO-US01850.
PF 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
FA Cheever MA, Hand-Zimmermann S;
XX MPI: 2001-476112/51.
DR N-PSDB; AAB23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.

XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 22; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASVCTGTDKRLPASPETHLDMRLHYQGGQVAVQNL 60
Db 1 MELAALCRWGLLALLPPGAASVCTGTDKRLPASPETHLDMRLHYQGGQVAVQNL 60
QY 61 ETTYPTNASLFLQDIOEVQGYVLIANQVRQVPLRLKATVKGTLQFEDNYALAVLDNG 120

Db 61 ETTYPTNASLFLQDIOEVQGYVLIANQVRQVPLRLKATVKGTLQFEDNYALAVLDNG 120
QY 121 DELNNTTVTAASPGGLRELOLSLTLELKGVLIQNPOLCYQDTLLMKDIFHKQNOLA 180
Db 121 DELNNTTVTAASPGGLRELOLSLTLELKGVLIQNPOLCYQDTLLMKDIFHKQNOLA 180
QY 181 LTLIDTNSRACHPSPCKSGKSCRMWSESSEDCSLTETVACAGCARCKGLPTDCCHQEC 240
Db 181 LTLIDTNSRACHPSPCKSGKSCRMWSESSEDCSLTETVACAGCARCKGLPTDCCHQEC 240
QY 241 AAGCTPRHSDCLACLFHNSGICELCPALVTYNTDTPESMPNBEGRYTFGASCVYACP 300
Db 241 AAGCTPRHSDCLACLFHNSGICELCPALVTYNTDTPESMPNBEGRYTFGASCVYACP 300
QY 301 YNYLSTDVSGCTLVCPPLHNOEVTAEDEGTQRCCKSKPCARVCYGLGMEHREYAVTSAN 360
Db 301 YNYLSTDVSGCTLVCPPLHNOEVTAEDEGTQRCCKSKPCARVCYGLGMEHREYAVTSAN 360
QY 361 ELEPAGCKKIFGSLAFUPESFGDPPASTAPLOPELOVFETLEITGYLYISAMPDLP 420
Db 361 IOEFAGCKKIFGSLAFUPESFGDPPASTAPLOPELOVFETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNTQVIRGRILHNGAYSLTLOGIGISWLGSLRELSGLALIHNTHLCPVHTV 480
Db 421 DLSVFQNTQVIRGRILHNGAYSLTLOGIGISWLGSLRELSGLALIHNTHLCPVHTV 480
QY 481 PWDOLFRRPHQALLHTARPEDEVGSLACHQICARHGCMGPGPTCVNCSOFLRQEC 540
Db 481 PWDOLFRRPHQALLHTARPEDEVGSLACHQICARHGCMGPGPTCVNCSOFLRQEC 540
QY 541 VEECRVLQGLPREVYNAHCLPCHRECOPOGNSVTCGPEADQCAAHKADPEFCVAC 600
Db 541 VEECRVLQGLPREVYNAHCLPCHRECOPOGNSVTCGPEADQCAAHKADPEFCVAC 600
QY 601 PSGVPRDLSTYMPKFPDEEGACPCPCINCHSCVDLDDGCPAECRASLTSTISAVG 660
Db 601 PSGVPRDLSTYMPKFPDEEGACPCPCINCHSCVDLDDGCPAECRASLTSTISAVG 660
QY 661 ILVVVLGVFGLIILKRQOKIRKYTRRLLOETELVEPLTPSGAMNQOMLILKEQYI 720
Db 661 ILVVVLGVFGLIILKRQOKIRKYTRRLLOETELVEPLTPSGAMNQOMLILKEQYI 720
QY 721 KANSKFIGITEL-TVYGIWIPDGENYKIPVAIKVIRENTSPANKELIDEAAYMAGVGS 779
Db 721 KANSKFIGITEL-TVYGIWIPDGENYKIPVAIKVIRENTSPANKELIDEAAYMAGVGS 779
QY 780 PYVSRLLGICLTSTVOLVTOLMPYGCILDVRENRGLSGODLLNMCQIAKMSYLEDV 839
Db 780 PYVSRLLGICLTSTVOLVTOLMPYGCILDVRENRGLSGODLLNMCQIAKMSYLEDV 839
QY 840 RLVRHDLAARNVLKSPNHVKITDPGLARLIDIDETRYHADGKVPIMMALLESILRRF 899
Db 840 RLVRHDLAARNVLKSPNHVKITDPGLARLIDIDETRYHADGKVPIMMALLESILRRF 899
QY 900 THQSDVMSYGVTTWELMTFGAKPYDGIIPABEIPDLLEKSGRLPQPPICITDVMYMKW 959
Db 900 THQSDVMSYGVTTWELMTFGAKPYDGIIPABEIPDLLEKSGRLPQPPICITDVMYMKW 959
QY 960 MIDSECRPRFELVSEFSRMAKRDQRFVIVQNEDLGPASPLDSTFTRSLLEDMDGDVLD 1019
Db 960 MIDSECRPRFELVSEFSRMAKRDQRFVIVQNEDLGPASPLDSTFTRSLLEDMDGDVLD 1019
QY 1020 AEEYLVPQGGFFCCDPAPAGAGVHHRSSSTRSGGGDLTLGLEPSEEARPRPLAPSE 1079
Db 1020 AEEYLVPQGGFFCCDPAPAGAGVHHRSSSTRSGGGDLTLGLEPSEEARPRPLAPSE 1079
QY 1080 GAGSDVFDGDLGMGAAGKIQSLPTHDSPLQRYSDPTVPLPSETDGVAPLTGSPQPEY 1139
Db 1080 GAGSDVFDGDLGMGAAGKIQSLPTHDSPLQRYSDPTVPLPSETDGVAPLTGSPQPEY 1139
QY 1140 VAGQDVPRPQPSRPGSLPAARPAATLERAKTSLPGKNGVYKDFAFGAVENPEYLT 1199

DB 1140 VNQPDVRFPPSPREGGLPAPRPAAGATLEPRKTLSPCKNGVXKDVAFGAVENPEYLTP 1199
QY 1200 QGGAAPQHPHPAPSPAFDNLYYWDQDPPERGA PPSFGKSTPTAENPEYLGLDVPV 1255
DB 1200 QGGAAPQHPHPAPSPAFDNLYYWDQDPPERGA PPSFGKSTPTAENPEYLGLDVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
AC AAG88267;
XX 11-SEP-2001 (first entry)
DT 11-SEP-2001 (first entry)
XX HER2/neu amino acid sequence.
DE HER2/neu amino acid sequence.
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KM immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX Homo sapiens.
OS Homo sapiens.
XX MO200141787-A1.
PN 14-JUN-2001.
PD 11-DEC-2000; 2000MO-US33591.
PF 10-DEC-1999; 99US-0458299.
XX (EPRM-) EPIMUNE INC.
PA Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX WPI; 2001-374995/39.
DR An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
PS The present invention describes isolated prepared HER2/neu epitopes (1).
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX Sequence 1255 AA;
SQ

Query Match 97.4%; Score 6640; DB 22; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAALRKGILLALLPFGAASQVCTGDMKRLPASETHLMDLRHLYGQGVQVGNL 60
DB 1 MELAALRKGILLALLPFGAASQVCTGDMKRLPASETHLMDLRHLYGQGVQVGNL 60
QY 61 ELTYLPTNASLSFQDIOEVQGVYLLAHNQVQVPLQRIYRGVQLFEDNALAVLDNG 120
DB 61 ELTYLPTNASLSFQDIOEVQGVYLLAHNQVQVPLQRIYRGVQLFEDNALAVLDNG 120
QY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLTQRPOLCYODITLWKDIFHKNNQLA 180
DB 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLTQRPOLCYODITLWKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPCKGSGCWSESSDDQSLRTYVACAGCARCKGPLEPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPCKGSGCWSESSDDQSLRTYVACAGCARCKGPLEPTDCHEQC 240
QY 241 AAGCTGPKSDCLACHFNHSGICEIHCALVTYNTDTFESMPNEGRTFGASCVTACP 300
DB 241 AAGCTGPKSDCLACHFNHSGICEIHCALVTYNTDTFESMPNEGRTFGASCVTACP 300
QY 301 YNTLSTDVSGCTIVCPAHQEVTAEDGTQRCCKSPCARVCYGLMOYIKANSFKIGIT 360
DB 301 YNTLSTDVSGCTIVCPAHQEVTAEDGTQRCCKSPCARVCYGLMEHLREVRVTSAN 360
QY 361 ELFPAGCKKIFGSLATPSPFGDDPASNAPLOPQLOVFEIETITGYLYSAMPDSLP 420
DB 361 IQEFAQCKKIFGSLATPSPFGDDPASNAPLOPQLOVFEIETITGYLYSAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRELSSGLALIHNNHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRELSSGLALIHNNHLCFVHTV 480
QY 481 PMQULFRNHQALHTANRPEDECEYGEGLAQHQLCARHGWCPGPTQCVCNSQFLRGEC 540
DB 481 PMQULFRNHQALHTANRPEDECEYGEGLAQHQLCARHGWCPGPTQCVCNSQFLRGEC 540
QY 541 VEECRVLOGLPREYVYARHCLPCHHECOPONGSVTCFPEADQCAVCHYVDPFCVARC 600
DB 541 VEECRVLOGLPREYVYARHCLPCHHECOPONGSVTCFPEADQCAVCHYVDPFCVARC 600
QY 601 PGVCPDLSYPMIRKPEDEGACQCPINCTHSCVDLDDKCPAEOBASPLTISAVG 660
DB 601 PGVCPDLSYPMIRKPEDEGACQCPINCTHSCVDLDDKCPAEOBASPLTISAVG 660
QY 661 ILLVYVLAGVFGILLKRBQOKIRKXTMRLLQETELVEPLTPSGAMPQAOAMRIKETEL 720
DB 661 ILLVYVLAGVFGILLKRBQOKIRKXTMRLLQETELVEPLTPSGAMPQAOAMRIKETEL 720
QY 721 KANSKFIGITEL-TVYKGIWIPDGENVAKPAIKVLRNTPSKANKEILDEAYVMAGVS 779
DB 721 KANSKFIGITEL-TVYKGIWIPDGENVAKPAIKVLRNTPSKANKEILDEAYVMAGVS 779
QY 780 EYVSLLLGLCLTSTYQVLTQMLPQCGCLLDHRENRGRIGSDDLNMCOIAKGSYLEDV 839
DB 780 EYVSLLLGLCLTSTYQVLTQMLPQCGCLLDHRENRGRIGSDDLNMCOIAKGSYLEDV 839
QY 840 RLVHRDLAARVLYVSPNHVKITPFGARLLDIDETETHADGGVPIKMMALSELRRRF 899
DB 840 RLVHRDLAARVLYVSPNHVKITPFGARLLDIDETETHADGGVPIKMMALSELRRRF 899
QY 900 THQSVMYSYGVTELMFTGAKPYDGPAREIPDLKGERLPQPICTIDVYIMWKCM 959
DB 900 THQSVMYSYGVTELMFTGAKPYDGPAREIPDLKGERLPQPICTIDVYIMWKCM 959
QY 960 MIDSECRFPREIYSESRMARDDQRFVYVONEDLGASPLDSTFFYSLLIEDDMGDLVD 1019
DB 960 MIDSECRFPREIYSESRMARDDQRFVYVONEDLGASPLDSTFFYSLLIEDDMGDLVD 1019
QY 1020 AEEYIVPQGFCCDPAPAGAGVWHHRSSSTRSGGDDTLTGIEPSEEAAPSPAPASE 1079
DB 1020 AEEYIVPQGFCCDPAPAGAGVWHHRSSSTRSGGDDTLTGIEPSEEAAPSPAPASE 1079

DY 1 MELALCRWGLLALLPPGAASQVCTGTDMKRLPASETHLDMLRHLYGCCVVQGNT 60

DB 1 MELALCRWGILLALLPPGAASQVCTGTDMKRLPASETHLDMLRHLYGCCVVQGNIT 60

QY 1080 GAGSDVDFDGLGNGAKGLQLPLTHDPELQRSEDPVPLPSETDGYAAPLTSPQPEY 1139

Db 1080 GAGSDVDFDGLGNGAKGLQLPLTHDPELQRSEDPVPLPSETDGYAAPLTSPQPEY 1139

Db 1140 VNQPDVAFPPSPREGFLPAAPAGATLEPPKTLSPGKGVAVDVAFGAAVENEYLLP 1199
QY 1200 QGGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSPSTFKGTPAENPEYLGLDVEV 1255
Db 1200 QGGAAFPQHPPPAFSPAFDNLVYWDODPPERGAPSPSTFKGTPAENPEYLGLDVEV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B*44-restricted epitope"
PN MO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
PA Hand-zimmerman S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI Mcneill PD, Vedvick TS;
PI MPI: 2002-280758/32.
DR N-PSDB; AAD02743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX
XX Disclosure; Page 114-117; 129pp; English.
XX
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B*44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer.
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridization, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 23; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAAICRMWLLALLPRGAASQVCTGDMKRLRLPASEETHLDMRLHYGCCVQVQNL 60
Db 1 MELAAICRMWLLALLPRGAASQVCTGDMKRLRLPASEETHLDMRLHYGCCVQVQNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHQVQVPLQRIIRIRGTQLFEDNALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHQVQVPLQRIIRIRGTQLFEDNALAVLNG 120
QY 121 DPLNNTTPVTGASPGGLRELQLRLSTELLKGVLIQRNPOLCYQDTILMKDI FHNQNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQLRLSTELLKGVLIQRNPOLCYQDTILMKDI FHNQNQLA 180
QY 181 LTIIDNRSRACHPCSPMGKSGRCWSESESDQSILTRIVCAGGACRCGPLETDCCHQC 240
Db 181 LTIIDNRSRACHPCSPMGKSGRCWSESESDQSILTRIVCAGGACRCGPLETDCCHQC 240
QY 241 AAGCTGPKASDCLACLFHNSGICEIHCPLATVYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCTGPKASDCLACLFHNSGICEIHCPLATVYNTDFESMPNDEGRYTFGASCVTACP 300
QY 301 YNLTSTDVSSCTLVCPILHNGEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNLTSTDVSSCTLVCPILHNGEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLESPFDGDSANTAPLQPPQLQVFTELEITGYLIAMPSLSP 420
Db 361 IQEFAGCKKIFGSLAFLESPFDGDSANTAPLQPPQLQVFTELEITGYLIAMPSLSP 420
QY 421 DLSVFQNLQVIRGILIHNGAYSLLTQGLGISWLGRLSRLREGSGIALIHNTHTLCFVATV 480
Db 421 DLSVFQNLQVIRGILIHNGAYSLLTQGLGISWLGRLSRLREGSGIALIHNTHTLCFVATV 480
QY 481 PMDQLFRRPHQALLHTANRPEDECEGEGLAGHQLCARHCGPPTQCVNCSQPLRGQEC 540
Db 481 PMDQLFRRPHQALLHTANRPEDECEGEGLAGHQLCARHCGPPTQCVNCSQPLRGQEC 540
QY 541 VEECRVQLQPREYVNAHCLPCHBECOPQNGSVTCFPEADQCVACAHYCDPFCVARC 600
Db 541 VEECRVQLQPREYVNAHCLPCHBECOPQNGSVTCFPEADQCVACAHYCDPFCVARC 600
QY 601 PSGVPPDLSYPIKMFPPDEBAGCQPCPINCCHSCVDLDDKCPAEPASPLTISAVVG 660
Db 601 PSGVPPDLSYPIKMFPPDEBAGCQPCPINCCHSCVDLDDKCPAEPASPLTISAVVG 660
QY 661 ILVVVLGVFGIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAMRILKEQYI 720
Db 661 ILVVVLGVFGIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAMRILKEQYI 720
QY 721 KANSKFIGITEL-TYKGIWIPDGENYKIPVAIKVLBENTSPKANKETLDEAYVMAGVGS 779
Db 721 KANSKFIGITEL-TYKGIWIPDGENYKIPVAIKVLBENTSPKANKETLDEAYVMAGVGS 779
QY 780 PVSLLIGICLTSTYQVLTQMPYGCGLDHYRENRGSLGSDLLNMCQIAKMGSYLEDV 839
Db 780 PVSLLIGICLTSTYQVLTQMPYGCGLDHYRENRGSLGSDLLNMCQIAKMGSYLEDV 839
QY 840 RLVHRLAARNVLKSPNHYKIDTFGLARLLIDETVYHADGGVPIKMMALSLIRRRF 899
Db 840 RLVHRLAARNVLKSPNHYKIDTFGLARLLIDETVYHADGGVPIKMMALSLIRRRF 899
QY 900 THQSDVMSYGTVWELMTFGAKPYDGIIPABEIPDLLEKGERLQPPICITIDVYMIWKW 959
Db 900 THQSDVMSYGTVWELMTFGAKPYDGIIPABEIPDLLEKGERLQPPICITIDVYMIWKW 959
QY 960 MIDSECRPFRELIVSEFSRMAKRDQRFVIVQNEIDLPASPLDSTFVRSILEDMDMDVDV 1019
Db 960 MIDSECRPFRELIVSEFSRMAKRDQRFVIVQNEIDLPASPLDSTFVRSILEDMDMDVDV 1019
QY 1020 ABEYLVVQQGFFCCDDPAPGAGGVVHHRHSSSTRSGGDLTLLGPESEEAAPSPLAPSE 1079
Db 1020 ABEYLVVQQGFFCCDDPAPGAGGVVHHRHSSSTRSGGDLTLLGPESEEAAPSPLAPSE 1079

Db 1020 AEEYLVPGQGFCDPAPAGAGWVHRHRSSTNRSGGDLTGLERSEERAPSLADSE 1079
 Qy 1080 GAGSDVDPDGLGMAKAGLQSLTTHPSPLQRYSEDPYPLBSETDGYVAPLTCSPOPEY 1139
 Db 1080 GAGSDVDFDGLGMAKAGLQSLTTHPSPLQRYSEDPYPLBSETDGYVAPLTCSPOPEY 1139
 Qy 1140 VNQDVPRPQPSREGLPAPARAGATLEBAKTLSSGKGVKDVAFGGAIVENPEYLTLP 1199
 Db 1140 VNQDVPRPQPSREGLPAPARAGATLEBAKTLSSGKGVKDVAFGGAIVENPEYLTLP 1199
 Qy 1200 QGGAAPQPPPPAFSPAFDNLVWDQDPPERGAPESTFKGTPTAENPEYLGADV 1255
 Db 1200 QGGAAPQPPPPAFSPAFDNLVWDQDPPERGAPESTFKGTPTAENPEYLGADV 1255
 RESULT 13
 AAMS1143
 ID AAMS1143 standard; Protein; 1255 AA.
 AC AAMS1143;
 XX 17-JUN-2002 (first entry)
 DT 17-JUN-2002 (first entry)
 DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
 XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KM Tyrosine kinase; receptor; c-erbB2; gene therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 1..653 "extracellular domain"
 FT Domain 676..1255
 FT Domain /note= "intracellular domain"
 FT Domain 990..1255
 FT Domain /note= "phosphorylation domain"
 XX
 PN MO200212341-A2.
 XX
 PD 14-FEB-2002.
 PF 03-AUG-2001; 2001WO-US24283.
 PR 03-AUG-2000; 2000US-0632507.
 PA (CORI-) CORIXA CORP.
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Cheever MA, Gheysen D;
 DR WPI: 2002-241743/29.
 DR N-PSDB; ABR92250.
 XX
 PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain -
 PS Claim 68; Fig 7; 141pp; English.
 XX
 CC The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its Delta fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.
 CC
 XX
 SQ Sequence 1255 AA;
 Query Match 97.4%; Score 6640; DB 23; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;
 Qy 1 MELAALCRWGLLALLPQGAATQVCTGDMKRLPASPETHLDMRLHYGQCVVQGNL 60
 Db 1 MELAALCRWGLLALLPQGAATQVCTGDMKRLPASPETHLDMRLHYGQCVVQGNL 60
 Qy 61 ELTYLPTNASLFLDIOEVQCYVLIANNQVRYVPLQRLIRVGTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLDIOEVQCYVLIANNQVRYVPLQRLIRVGTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVYGASPGGLRELOLRSTLTKGGLVLIORNPOLCYODTLMKDIFHKNNOLA 180
 Db 121 DPLNNTPTVYGASPGGLRELOLRSTLTKGGLVLIORNPOLCYODTLMKDIFHKNNOLA 180
 Qy 181 LTLIDTNSRACHPCSPCKGSRGMBESSEDCQSLRTVACGACARCKGLPTDCHEQC 240
 Db 181 LTLIDTNSRACHPCSPCKGSRGMBESSEDCQSLRTVACGACARCKGLPTDCHEQC 240
 Qy 241 AAGCTGPHSDCLGHNHSGIEMHCPALVYNTDTFESMPBERYTFGASCVTACP 300
 Db 241 AAGCTGPHSDCLGHNHSGIEMHCPALVYNTDTFESMPBERYTFGASCVTACP 300
 Qy 301 YNYLSTDVSGCTLVCPHNGEVTAEEDTGRCCKSKCARVCYGLGMQYIKANSKFIQIT 360
 Db 301 YNYLSTDVSGCTLVCPHNGEVTAEEDTGRCCKSKCARVCYGLGMQYIKANSKFIQIT 360
 Qy 361 ELEFAGCKKIFGSLAFLEPFDGPPASNTAPLOEQVETLEETITGYLYISAMPDSL 420
 Db 361 IOEFAGCKKIFGSLAFLEPFDGPPASNTAPLOEQVETLEETITGYLYISAMPDSL 420
 Qy 421 DLSVFQNLQVIRGRILNNGAVSLTQGLGSMGLRSLRGLALIHNTLTCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILNNGAVSLTQGLGSMGLRSLRGLALIHNTLTCFVHTV 480
 Qy 481 PWDOLFNRPHQALHTANRPDECEVSGGLACHQLCARGHCGMPGCPVNCQSLFNGQEC 540
 Db 481 PWDOLFNRPHQALHTANRPDECEVSGGLACHQLCARGHCGMPGCPVNCQSLFNGQEC 540
 Qy 541 VEEGRVLOGLPREVYVNRHCLCPHPEQOPNGSVTCGPADQCVACAHYKBPFCVANC 600
 Db 541 VEEGRVLOGLPREVYVNRHCLCPHPEQOPNGSVTCGPADQCVACAHYKBPFCVANC 600
 Qy 601 PSQVKPLSTWPIWKFDEBGAQCPGICNTHSCVDLDKGCAPAEORASPLTISVAVG 660
 Db 601 PSQVKPLSTWPIWKFDEBGAQCPGICNTHSCVDLDKGCAPAEORASPLTISVAVG 660
 Qy 661 ILVVLVGVVFGILIRKROQKIRKYMRLDETELVEPLTSGAMPNQAQRILKEQYI 720
 Db 661 ILVVLVGVVFGILIRKROQKIRKYMRLDETELVEPLTSGAMPNQAQRILKEQYI 720
 Qy 721 KANSKFIQITEL-TVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAGVS 779
 Db 721 R-KYKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAGVS 779

QY 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRLGSDLLNMCMQIAKMSYLEDV 839
 Db 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRLGSDLLNMCMQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRRRF 899
 Db 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRRRF 899
 QY 900 THQSDVWSYGVTVLMTFGAKPYDGI:PAEITDULEKGBRLPQPICTIDVYMMVNCW 959
 Db 900 THQSDVWSYGVTVLMTFGAKPYDGI:PAEITDULEKGBRLPQPICTIDVYMMVNCW 959
 QY 960 MIDSECRPRELSEFSRMAADPQRFVVIQNEDELSPASPLDSTFYRSLLEDDMDGLVD 1019
 Db 960 MIDSECRPRELSEFSRMAADPQRFVVIQNEDELSPASPLDSTFYRSLLEDDMDGLVD 1019
 QY 1020 AEEYLYPQGGFFCDDPAPGAGMWHHRSSSTRSGGDLTLGLESESEARSLASE 1079
 Db 1020 AEEYLYPQGGFFCDDPAPGAGMWHHRSSSTRSGGDLTLGLESESEARSLASE 1079
 QY 1080 GAGSDVFDGDLGMAAGKQSLPPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 Db 1080 GAGSDVFDGDLGMAAGKQSLPPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1140 VNQPDVRRQPSREGEPLPARPAGATLERAKTSLSGKNGVYKDVAFAGAVENBEYLTP 1139
 Db 1140 VNQPDVRRQPSREGEPLPARPAGATLERAKTSLSGKNGVYKDVAFAGAVENBEYLTP 1139
 QY 1200 QGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPESTFGKPTPAENBEYLGLDVPV 1255
 Db 1200 QGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPESTFGKPTPAENBEYLGLDVPV 1255

RESULT 14
 AAU77114
 ID AAU77114 standard; Protein; 1255 AA.
 AC AAU77114;
 DT 05-JUN-2002 (first entry)
 DE Human Her-2/neu polypeptide.
 KW Human; Her-2/neu; cytostatic; haematological malignancy; CML; acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL; chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS; Hodgkin's lymphoma; T cell therapy.
 OS Homo sapiens.
 FN WO200213847-A2.
 PD 21-FEB-2002.
 PF 13-AUG-2001; 2001WO-US25408.
 PR 14-AUG-2000; 2000US-0638280.
 PS 28-SEP-2000; 2000US-0675904.
 PA (CORI-) CORIXA CORP.
 PI Gaiger A, Cheever MA, Hand-zimmermann S;
 DR WPI; 2002-280741/32.
 DB N-PSDB; ABK10730.
 PT Inhibiting haematological malignancy development by administering polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide encoding the polypeptide, or antigen presenting cells expressing the polypeptide -
 PS Disclosure; Page 71-74; 74pp; English.

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX

Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 23; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAAICRKGILLALPPGAASTVCTGTDKMLRLPASPEHLDMLNHLVQGCQVVOGSL 60
 Db 1 MELAAICRKGILLALPPGAASTVCTGTDKMLRLPASPEHLDMLNHLVQGCQVVOGSL 60
 QY 61 ELTYLPTNASLFLQDIOEVQGYLIAHNVQVPLQRLRIVRGTOVFEENYALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIOEVQGYLIAHNVQVPLQRLRIVRGTOVFEENYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELOLRSTELIKGGVLIQNPOLCTYDITLWMDIFHKNNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELOLRSTELIKGGVLIQNPOLCTYDITLWMDIFHKNNQLA 180
 QY 181 LFLIDPNSRAHPSPKCKSRCKGSESSDCCGLTRTVACAGGARCKGLPDDCCEQC 240
 Db 181 LFLIDPNSRAHPSPKCKSRCKGSESSDCCGLTRTVACAGGARCKGLPDDCCEQC 240
 QY 241 AAGCTGPKXSDCLAFHNSGICELHCPALVTYNTDFESMPNPEGRYTFGASCYTACP 300
 Db 241 AAGCTGPKXSDCLAFHNSGICELHCPALVTYNTDFESMPNPEGRYTFGASCYTACP 300
 QY 301 YNYLSTDVSGCTLVCPLENOEVTAEADGTORCEKSKCAVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTDVSGCTLVCPLENOEVTAEADGTORCEKSKCAVCYGLGMQYIKANSKFIGIT 360
 QY 361 ELEFAGCKKIFGSLAFLEBSFDGSPASNTABLOEQLOVEETLEITGYLYISAMPDSL 420
 Db 361 IOEFAGCKKIFGSLAFLEBSFDGSPASNTABLOEQLOVEETLEITGYLYISAMPDSL 420
 QY 421 DLSVQNLQVIRGRILHNGAYSLLTQGLGISWGLRLREUGSLALIHNTLCEVHTV 480
 Db 421 DLSVQNLQVIRGRILHNGAYSLLTQGLGISWGLRLREUGSLALIHNTLCEVHTV 480
 QY 481 PWDOLFENPHQALLHTANRPEDECVGEGLAHQLCARHGMGRPTQCVNCSQFLRGQEC 540
 Db 481 PWDOLFENPHQALLHTANRPEDECVGEGLAHQLCARHGMGRPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYNAHCLPCHPEQCPQNGSYTCGPRADQVCANHKDPFPCVAC 600
 Db 541 VEECRVLQGLPREYNAHCLPCHPEQCPQNGSYTCGPRADQVCANHKDPFPCVAC 600
 QY 601 PSQVPRDLSYPMWKFPEDEGACQPCPINTCHSCVDLDKCGPREGASPLSTIISAVVG 660
 Db 601 PSQVPRDLSYPMWKFPEDEGACQPCPINTCHSCVDLDKCGPREGASPLSTIISAVVG 660
 QY 661 ILLVVVLGVFGIILIKRQCKIRRYTMRRLLOETELVEPLTPSGAMPNQOMKILETEL 720
 Db 661 ILLVVVLGVFGIILIKRQCKIRRYTMRRLLOETELVEPLTPSGAMPNQOMKILETEL 720
 QY 721 KANSKFIGITEL-TVYKGIWIPDENVKIPVALIVRENTSPRANKELIDEAYVMAVGS 779
 Db 721 KANSKFIGITEL-TVYKGIWIPDENVKIPVALIVRENTSPRANKELIDEAYVMAVGS 779
 QY 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRLGSDLLNMCMQIAKMSYLEDV 839
 Db 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRLGSDLLNMCMQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRRRF 899
 Db 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRRRF 899

```

Db      840 RLVHRDLAARNVLYKSPNHNKXITDGLARLLDIDETEHADGGKVPKXMALESILRRRF 899
QY      900 THOSDWSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
Db      900 THOSDWSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
QY      960 MIDSECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
Db      960 MIDSECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
QY      1020 AEEYLVPOQGFCCPDPAAGAGMWHHRSSSTSGGDLTLGLEPSEEEAPRSPPLAPSE 1079
Db      1020 AEEYLVPOQGFCCPDPAAGAGMWHHRSSSTSGGDLTLGLEPSEEEAPRSPPLAPSE 1079
QY      1080 GAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
Db      1080 GAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
QY      1140 VNQDPVRPQPPSPREGRPLPARPAGATLERAKTILSPKNGVYKDVPAFGAVENPEYLTLP 1199
Db      1140 VNQDPVRPQPPSPREGRPLPARPAGATLERAKTILSPKNGVYKDVPAFGAVENPEYLTLP 1199
QY      1200 QGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAPSPSTFKGTPTAENPEYLGIDVPY 1255
Db      1200 QGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAPSPSTFKGTPTAENPEYLGIDVPY 1255

RESULT 15
AAR39568 standard; Protein; 1433 AA.
AC      AAR39568;
DT      07-FEB-1994 (first entry)
DE      Sequence of c-erbB-2 tumour antigen.
KW      Tumour antigen; c-erbB-2; glycoprotein.
OS      Homo sapiens.
PN      M09316185-A.
PD      19-AUG-1993.
PF      05-FEB-1993; 93MO-US01055.
PR      06-FEB-1992; 92US-0831967.
XX      (CETU) CETUS ONCOLOGY CORP.
PA      (CREA-) CREATIVE BIOMOLECULES INC.
PI      Houston TX, Huston JS, Oppermann H, Ring DB;
XX      WPI; 1993-272869/34.
DR      N-PSDB; AAQ46083.
XX      New single chain Fv polypeptide binding to C-erbB-2 tumour
PT      antigen - for imaging or treating breast or ovarian cancer etc.
XX      Disclosure; pages 48-54; 87pp; English.
PS      CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC      tumour cells, such as breast and ovarian tumour cells, which is an
CC      approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC      pt. of about 5.3 (see AAQ46083; AAR39568). X in AAR39568 represents
CC      the location of a stop codon in AAQ46083.
XX      Sequence 1433 AA;

```

```

Query Match 96.8%; Score 6597; DB 14; Length 1433;
Best Local Similarity 97.0%; Pred. No. 0;

Matches 1218; Conservative 10; Mismatches 26; Indels 2; Gaps 2
QY      1 MELALCRWGILLALLPPGAASIQVCTGDMKRLRSPETHLDMRLHYQGVQVQNL 60
Db      1 MELALCRWGILLALLPPGAASIQVCTGDMKRLRSPETHLDMRLHYQGVQVQNL 60
QY      61 ELTYLPTNASLFIQDIQEVQYVLLIAHNOVRQVPLQRLIRVAGTOLFEDNYLALVDNG 120
Db      61 ELTYLPTNASLFIQDIQEVQYVLLIAHNOVRQVPLQRLIRVAGTOLFEDNYLALVDNG 120
QY      121 DPLNNTTPVTGASPGGLREQLRSLTEILKGVYLIQNPQLCYQDITLWKDIFHXNQLA 180
Db      121 DPLNNTTPVTGASPGGLREQLRSLTEILKGVYLIQNPQLCYQDITLWKDIFHXNQLA 180
QY      181 LTLIDNRSRACHSCSMWGSCKSSCKWESSSDQSLRTYVACGACACKGPLPDDCCEOC 240
Db      181 LTLIDNRSRACHSCSMWGSCKSSCKWESSSDQSLRTYVACGACACKGPLPDDCCEOC 240
QY      241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGASCTYACP 300
Db      241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGASCTYACP 300
QY      301 YNYLSTVGSCTLYVCPILHNOEVTAEQTCRCKSPCARVCYGLQMOYIKANSKFIQIT 360
Db      301 YNYLSTVGSCTLYVCPILHNOEVTAEQTCRCKSPCARVCYGLQMOYIKANSKFIQIT 360
QY      361 ELFEAGCKKIFGSLAFPLPESFDDPDPSANTAPLOPELOVETLEETGYLYXISAMPDILP 420
Db      361 IQEPAQCKRIFGSLAFPLPESFDDPDPSANTAPLOPELOVETLEETGYLYXISAMPDILP 420
QY      421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELSGLALIHNNHLSFVHV 480
Db      421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELSGLALIHNNHLSFVHV 480
QY      481 PMDQLFENPHQALLHNTANREDECVGEGLACHQLCARHGCKMGPPIQVNCQFRLQGC 540
Db      481 PMDQLFENPHQALLHNTANREDECVGEGLACHQLCARHGCKMGPPIQVNCQFRLQGC 540
QY      541 VEEGRVLYGLPREYVNASHCLEPHPCQCPQNSVTCFGEBAQCACAHYKDPFCVARC 600
Db      541 VEEGRVLYGLPREYVNASHCLEPHPCQCPQNSVTCFGEBAQCACAHYKDPFCVARC 600
QY      601 PSQYKPDLSYMPIMKPEPDEGACQCPICNCTHSCVDLDKCCPAEQRASPLTISAVVG 660
Db      601 PSQYKPDLSYMPIMKPEPDEGACQCPICNCTHSCVDLDKCCPAEQRASPLTISAVVG 660
QY      661 ILVVVLYGVVFGILIKRPOOKIRXYTMRRLQETELVEPLTSSGAMPNOACMRILKEQYI 720
Db      661 ILVVVLYGVVFGILIKRPOOKIRXYTMRRLQETELVEPLTSSGAMPNOACMRILKEQYI 720
QY      721 KANSKFIQTTEL-TYVKGIMIPDGENVKIPVAKYLRENTSPKANKELIDEAYVAVAGVS 779
Db      721 KANSKFIQTTEL-TYVKGIMIPDGENVKIPVAKYLRENTSPKANKELIDEAYVAVAGVS 779
QY      780 PYVSRLLGICLSTVQVLTQMLPYGGLDHYENGRIGSODLLMWCMQIAKGSYLEDV 839
Db      780 PYVSRLLGICLSTVQVLTQMLPYGGLDHYENGRIGSODLLMWCMQIAKGSYLEDV 839
QY      840 RLVHRDLAARNVLYKSPNHNKXITDGLARLLDIDETEHADGGKVPKXMALESILRRRF 899
Db      840 RLVHRDLAARNVLYKSPNHNKXITDGLARLLDIDETEHADGGKVPKXMALESILRRRF 899
QY      900 THOSDWSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
Db      900 THOSDWSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
QY      960 MIDSECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
Db      960 MIDSECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
QY      1020 AEEYLVPOQGFCCPDPAAGAGMWHHRSSSTSGGDLTLGLEPSEEEAPRSPPLAPSE 1079
Db      1020 AEEYLVPOQGFCCPDPAAGAGMWHHRSSSTSGGDLTLGLEPSEEEAPRSPPLAPSE 1079

```

```
OY 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPYVPLPSETDGYVAPLTCSPQPEY 1139
Db 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPYVPLPSETDGYVAPLTCSPQPEY 1139
OY 1140 VNQPDVVRPQPPSPREGFLPAARPAAGATLERAKTILSPGKNGVVKDYFAFGAVENPEYLTP 1199
Db 1140 VNQPDVVRPQPPSPREGFLPAARPAAGATLERAKTILSPGKNGVVKDYFAFGAVENPEYLTP 1199
OY 1200 QGGAAPQPHPPPAFPAFDNLYWDDPPERGAAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1200 QGGAAPQPHPPPAFPAFDNLYWDDPPERGAAPSTFKGTPTAENPEYLGIDVPV 1255
```

Search completed: July 22, 2003, 09:16:44
Job time : 42.939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5741 Seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-653-675-12

Perfect score: 6776

Sequence: 1 MELALCRWGLLLALPPGA.....TFKGTPTANPEYLGUDPEV 1247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6602	97.4	1255	1	ERR2_HUMAN
2	5815	85.8	1257	1	ERR2_RAT
3	5799.5	85.6	1254	1	ERR2_MESAU
4	3064	45.2	1210	1	EGFR_HUMAN
5	3033	44.8	1210	1	EGFR_MOUSE
6	2915.5	43.0	1308	1	ERR4_HUMAN
7	2896	42.7	1308	1	ERR4_RAT
8	2612.5	38.6	1167	1	XRK_XIPMA
9	2390.5	35.3	1342	1	ERR3_HUMAN
10	2316.5	34.2	1339	1	ERR3_RAT
11	1935	28.6	1426	1	EGFR_DROME
12	1666.5	24.6	634	1	ERRB_ALV
13	1620	23.9	604	1	ERRB_AVIER
14	1547	22.8	540	1	ERRB_AVIER
15	1512	22.3	703	1	EGFR_CHICK
16	1287	19.0	1323	1	LT23_CAEL
17	1142.5	16.9	245	1	ERR2_MOUSE
18	725	10.8	1363	1	ERR3_MOUSE
19	708	10.4	1300	1	ERR3_MOUSE
20	704	10.4	1383	1	INSR_RAT
21	703.5	10.4	1372	1	INSR_MOUSE
22	699	10.3	1382	1	INSR_HUMAN
23	697	10.3	1297	1	INSR_HUMAN
24	696.5	10.3	1300	1	INSR_HUMAN
25	688	10.2	1607	1	ERR3_MOUSE
26	682	10.1	1477	1	ERR3_MOUSE
27	648	9.6	1367	1	ERR3_MOUSE
28	635	9.4	1373	1	ERR3_MOUSE
29	631.5	9.3	1370	1	ERR3_MOUSE
30	618	9.1	1390	1	ERR3_MOUSE
31	610	9.0	2146	1	ERR3_MOUSE
32	593	8.8	1114	1	ERR3_MOUSE
33	591	8.7	987	1	ERR3_MOUSE

34	587.5	8.7	984	1	ERR3_MOUSE
35	580.5	8.6	984	1	ERR3_MOUSE
36	574.5	8.5	984	1	ERR3_MOUSE
37	574	8.5	987	1	ERR3_MOUSE
38	571	8.4	757	1	ERR3_MOUSE
39	571	8.4	902	1	ERR3_MOUSE
40	569	8.4	1068	1	ERR3_MOUSE
41	566.5	8.4	1053	1	ERR3_MOUSE
42	565.5	8.3	977	1	ERR3_MOUSE
43	563	8.3	1052	1	ERR3_MOUSE
44	561	8.3	1055	1	ERR3_MOUSE
45	560.5	8.3	985	1	ERR3_MOUSE

ALIGNMENTS

RESULT 1
ERR2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Ref. 05, Created)
DT 13-AUG-1987 (Ref. 05, Last sequence update)
DT 15-JUN-2002 (Ref. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
SA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.",
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=66070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.",
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=66016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RA "A v-erbB-related growth factor receptor, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.",
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.",
RL Genomics 15:426-429(1993).
RN [5]
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN ALPHA AND AMPHIREGULIN.
RN CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
RN (POTENTIAL).
RN SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.783; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M11767; AAA35808.1; -
DR EMBL: M11761; AAA35808.1; JOINED.
DR EMBL: M11762; AAA35808.1; JOINED.
DR EMBL: M11763; AAA35808.1; JOINED.
DR EMBL: M11764; AAA35808.1; JOINED.
DR EMBL: M11765; AAA35808.1; JOINED.
DR EMBL: M11766; AAA35808.1; JOINED.
DR EMBL: M11730; AAA75493.1; -
DR EMBL: M12036; AAA35978.1; -
DR EMBL: X03363; CAA27060.1; -
DR PIR: A25491; A25491.
DR PIR: A24571; A24571.
DR HSSP: P11362; 1FGK.
DR Genew: HGNC:3430; ERBB2.
DR MIM: 164870; -
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Polymorphism.
KW
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 227 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 BY SIMILARITY.
FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 654 654 I->V.
FT VARIANT 655 655 P->V.
FT VARIANT 655 655 I->V.
FT CONFLICT 1170 1170 P->A (IN REF. 2).
FT SEQUENCE 1255 AA; 137909 MM; 39BDPDA04DCP962 CRC64;
SQ

Query Match 97.4%; Score 6602; DB 1; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps

QY 1 MELAAACRWGLLALLPFGASTQVCTGTDMKRLPASPEFTHLMLRHLVGGCOVQONT 60
DB 1 MELAAACRWGLLALLPFGASTQVCTGTDMKRLPASPEFTHLMLRHLVGGCOVQONT 60
QY 61 ELTYLPTNASLFLQDIEVQGVYIAHNOYVPLQSLVRGTOUFEDNYALAVLDNG 121
DB 61 ELTYLPTNASLFLQDIEVQGVYIAHNOYVPLQSLVRGTOUFEDNYALAVLDNG 121
QY 121 DELNNTPTTGASPGGLRELQSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 181
DB 121 DELNNTPTTGASPGGLRELQSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 181
QY 181 LTLIDNRSRACHPSPCKSRCKSGESSECCSLTITVCAAGGARCKGPLPTDCCHEOC 241
DB 181 LTLIDNRSRACHPSPCKSRCKSGESSECCSLTITVCAAGGARCKGPLPTDCCHEOC 241
QY 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTTFESVPNPEGRYTFGASCYACP 301
DB 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTTFESVPNPEGRYTFGASCYACP 301
QY 301 VNYLSTVQSGCTLVCPHLNQGVTAEQDQRECKSKPCARVCYGLGKQVIKANSKFIQIT 361
DB 301 VNYLSTVQSGCTLVCPHLNQGVTAEQDQRECKSKPCARVCYGLGKQVIKANSKFIQIT 361
QY 361 ELEFAGCKKIFGSLAFPESEFDGPASNTAPLOPEQVFEITBEITGYIYSAMPDLP 421
DB 361 ELEFAGCKKIFGSLAFPESEFDGPASNTAPLOPEQVFEITBEITGYIYSAMPDLP 421
QY 421 DLVSQNLQVIRGILHNGAVSLTLOGLISWGLRSLRELSSGLALIHNTLCEVHTV 481
DB 421 DLVSQNLQVIRGILHNGAVSLTLOGLISWGLRSLRELSSGLALIHNTLCEVHTV 481
QY 481 PMDQFRPHALHTANRPDECVGEGGLACHOLCARGHGWGSGPQCVNCSQFLMGQSC 541
DB 481 PMDQFRPHALHTANRPDECVGEGGLACHOLCARGHGWGSGPQCVNCSQFLMGQSC 541
QY 541 VEECRVILQGLPREVYVNRHCLPCHEPCQPNQSVTFGPEADQCVACAHYKDPFPCVAC 601
DB 541 VEECRVILQGLPREVYVNRHCLPCHEPCQPNQSVTFGPEADQCVACAHYKDPFPCVAC 601
QY 601 PSQVAPLSTWPIKPFDEEGACQPCINTHSCVLDKGCABRASPPLYSAYVG 661
DB 601 PSQVAPLSTWPIKPFDEEGACQPCINTHSCVLDKGCABRASPPLYSAYVG 661
QY 661 ILTVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNOAQRIIKETEL 712
DB 661 ILTVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNOAQRIIKETEL 712
QY 712 ILTVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNOAQRIIKETEL 720
DB 712 ILTVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNOAQRIIKETEL 720

QY 713 RKVVLGSGAFGVYKGIWIPDGENVKTIPVAKVIRENTSPKANKELLIDENYVAGVSGP 772
 DB 721 RKVVLGSGAFGVYKGIWIPDGENVKTIPVAKVIRENTSPKANKELLIDENYVAGVSGP 780
 QY 773 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVENRRLGSOGLLWCMQIAKAGSYLEDVR 832
 DB 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVENRRLGSOGLLWCMQIAKAGSYLEDVR 840
 QY 833 LVHRDLAARNVLYKSNHVKITDPGLARLDIDETEHADGKVPIMWALESLRRFT 892
 DB 841 LVHRDLAARNVLYKSNHVKITDPGLARLDIDETEHADGKVPIMWALESLRRFT 900
 QY 893 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVXCM 952
 DB 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVXCM 960
 QY 953 IDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTYRSLLBEDDDMDGLVDA 1012
 DB 961 IDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTYRSLLBEDDDMDGLVDA 1020
 QY 1013 EBYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1072
 DB 1021 EBYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 QY 1073 AGSDVDFDGLGMAAAGLSLPTHDSPLQRYSEPTVPLSESDGYVAPLTGSPQREYV 1132
 DB 1081 AGSDVDFDGLGMAAAGLSLPTHDSPLQRYSEPTVPLSESDGYVAPLTGSPQREYV 1140
 QY 1133 NOPDVRPQPPRREGGLPAPAPAGATLERAKTISGKGVKVDVAFAGAVENREYVTPQ 1192
 DB 1141 NOPDVRPQPPRREGGLPAPAPAGATLERAKTISGKGVKVDVAFAGAVENREYVTPQ 1200
 QY 1193 GGAAPQHPPPAPSPFNDLWYWDQDPPRPGAPBETFGCTPAENPEYLGLDVPV 1247
 DB 1201 GGAAPQHPPPAPSPFNDLWYWDQDPPRPGAPBETFGCTPAENPEYLGLDVPV 1255
 RESULT 2
 ERB2_PAT
 ID ERB2_PAT STANDARD; PRT; 1257 AA.
 AC P06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (101867b2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).
 DE Receptor-related protein.
 CN ERB2 OR NEU.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=66118662; PubMed=3945311;
 RA Bargmann C.I., Hung W.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein.";
 RT Nature 319:226-230(1996).
 RN [2]
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=9122560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";
 RT Neuron 6:691-704(1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=9215181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofte F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";
 RT EMBO J. 11:43-48(1992).
 RL [1]
 CC FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE, GP130 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
 CC -1 CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC -1 TYROSINE PHOSPHATE
 CC -1 SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERB2 RECEPTORS.
 CC -1 THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1 PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).
 CC -1 SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X03362; CA227059.1; ALT_INIT.
 CC PIR: A24562; TVRNTU.
 CC HSP: P11362; IFGK.
 CC InterPro: IPR000494; EGFR_L domain.
 CC InterPro: IPR000719; Euk_Dkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_Pkinase.
 CC InterPro: IPR004019; YLP_motif.
 CC Pfam: PF00069; Pkinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF02757; YLP; 2.
 CC ProDom: PD000001; Euk_pkinase; 1.
 CC SMART: SM00261; Fu; 3.
 CC SMART: SM00219; Tyrc; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC KW Proto-oncogene; Disease mutation.
 CC SIGNAL
 FT CHAIN 1..21 POTENTIAL.
 FT 22..1257 RECEPTOR PROTEIN-TYROSINE KINASE ERB2-2.
 FT 22..654 EXTRACELLULAR (POTENTIAL).
 FT 655..677 POTENTIAL.
 FT DOMAIN 678..1257 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 159..369 CYS-RICH.
 FT DOMAIN 473..646 CYS-RICH.
 FT DOMAIN 722..989 PROTEIN KINASE.
 FT NP_BIND 728..736 ATP (BY SIMILARITY).
 FT BINDING 735..735 ATP (BY SIMILARITY).
 FT ACT_SITE 847..847 BY SIMILARITY.
 FT DISULFID 196..205 BY SIMILARITY.
 FT DISULFID 200..213 BY SIMILARITY.
 FT DISULFID 221..228 BY SIMILARITY.
 FT DISULFID 225..236 BY SIMILARITY.
 FT DISULFID 237..245 BY SIMILARITY.
 FT DISULFID 241..253 BY SIMILARITY.
 FT DISULFID 256..265 BY SIMILARITY.
 FT DISULFID 269..296 BY SIMILARITY.
 FT DISULFID 300..312 BY SIMILARITY.
 FT DISULFID 316..332 BY SIMILARITY.
 FT DISULFID 335..339 BY SIMILARITY.
 FT DISULFID 513..522 BY SIMILARITY.
 FT DISULFID 517..530 BY SIMILARITY.
 FT DISULFID 533..542 BY SIMILARITY.
 FT DISULFID 546..562 BY SIMILARITY.
 FT DISULFID 565..578 BY SIMILARITY.

FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RRS 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RRS 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MM; 6129264583011402 CRC64;

Query Match 85.8%; Score 5815; DB 1; Length 1257;
 Best Local Similarity 85.7%; Pred. No. 1,4e-304;
 Matches 1077; Conservative 55; Mismatches 115; Indels 10; Gaps 5;

QY 1 MELAALCRWGLLALLPFGAATQVCTGDMKLRUPASPEHLDMLRHLVQCQVVGQNL 60
 DB 1 MELAAMCRWGLLALLPFGIAGTQVCTGDMKLRUPASPEHLDMLRHLVQCQVVGQNL 60
 QY 61 ELTYLPNASTSLFLODIOEVQGVYVLAHQVQVPLQRIATRYGQLFEDNYALAVLDNG 120
 DB 61 ELTYVPANASTSLFLODIOEVQGVYVLAHQVQVPLQRIATRYGQLFEDNYALAVLDNR 120
 QY 121 DPLNNTPTV-GASPGALRELOLRSLTEILKGVLIQORNPOLCYODTILMKDIFKXNQL 179
 DB 121 DPQDNVAASIPRGTPRGELRELOLRSLTEILKGVLIQORNPOLCYODTILMKDIFKXNQL 180
 QY 180 ALLTIDTNSRAHPGSPMGCKSGRCWSESDQCSTLRTVCAAGCARKGKPLPTDCHEQ 239
 DB 181 APVDIDTNSRAHPGSPMGCKSGRCWSESDQCSTLRTVCAAGCARKGKPLPTDCHEQ 240
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVYNTDFEEMPRDEGRYFPGASCYVAC 299
 DB 241 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVYNTDFEEMPRDEGRYFPGASCYVAC 300
 QY 300 PNYVLTSDVGSCTLVCPHNOEYTAEDGTQRCCKSKPCARCYVGLGMOYIKANSKPIGI 359
 DB 301 PNYVLTSDVGSCTLVCPHNOEYTAEDGTQRCCKSKPCARCYVGLGMOYIKANSKPIGI 360
 QY 360 TELEPAGCKKIFGSLAFPLESPFGDPASTAPLOEQLOVFTLEITGYIKSWPDSL 419
 DB 361 NVOEFGCKKIFGSLAFPLESPFGDPASTAPLOEQLOVFTLEITGYIKSWPDSL 420
 QY 420 PDLSEFQNLQVIRGRILLHNGAYSLTLOGIGISWLGIRSLRELGSGLALIHNTLCEVHT 479
 DB 421 PDLSEFQNLQVIRGRILLHNGAYSLTLOGIGISWLGIRSLRELGSGLALIHNTLCEVHT 480
 QY 480 VPWDOLFRRPHOHLHTANRDEDE-CVSGGLACHOGLCARGHGWGPGPQCNCGOLFRLGQ 538
 DB 481 VPWDOLFRRPHOHLHTANRDEDE-CVSGGLACHOGLCARGHGWGPGPQCNCGOLFRLGQ 540
 QY 539 ECVEECRVLQGLFREYVNAHRLCPHPECOPONGSVTCFGPEADQCVACAHYKDPFCVA 598
 DB 541 ECVEECRVLQGLFREYVNAHRLCPHPECOPONGSVTCFGPEADQCVACAHYKDPFCVA 600
 QY 599 RCSSGVKPDLSYPIWKPFDESGACQPCINCTHSCVDLDKGCPCAGRASPLTISVAV 658
 DB 601 RCSSGVKPDLSYPIWKPFDESGACQPCINCTHSCVDLDKGCPCAGRASPLTISVAV 660
 QY 659 VGILLVVLGVVFGIIL---QYIKANS--KEIGITEL--PLTPSGAMPNOAQRILIKET 710
 DB 661 VGILLVVLGVVFGIILIKRRQKIRKYMRLLOSTELVEPLTISGAMPNOAQRILIKET 720
 QY 711 ELRKVYVLSGAFGYIYKGIWIPDEBNYKIPVALIVLEBNTSPANKKEIDEAIVMAVG 770
 DB 721 ELRKVYVLSGAFGYIYKGIWIPDEBNYKIPVALIVLEBNTSPANKKEIDEAIVMAVG 780
 QY 771 SPVSRLLGICLTSTYQVLTQMLPFGCLLDHRENRGLUGSODLLNMCMQIAKMSYLED 830

DB 781 SPVSRLLGICLTSTYQVLTQMLPFGCLLDHRENRGLUGSODLLNMCMQIAKMSYLED 840
 QY 831 VRLVRBDLAARNVLYKSPNHVKITDGLARLLIDDETEYHADGKVPIKMALESILRR 890
 DB 841 VRLVRBDLAARNVLYKSPNHVKITDGLARLLIDDETEYHADGKVPIKMALESILRR 900
 QY 891 FTQSGVMSYGVTLVMTFGKAPVDGIPAREIPLLKGEPLPPICTDVYMIWKC 950
 DB 901 FTQSGVMSYGVTLVMTFGKAPVDGIPAREIPLLKGEPLPPICTDVYMIWKC 960
 QY 951 WMIDSECRPRELVESEFARMARDQRFVIONEDLGPASPLDSTFTYSLLDDDMGDLV 1010
 DB 961 WMIDSECRPRELVESEFARMARDQRFVIONEDLGPASPLDSTFTYSLLDDDMGDLV 1020
 QY 1011 DAEEYLVPOQGFCCPDPAAGACGMHHRSSSTSSGCGDLTLGLEBEEAPSPPLAPS 1070
 DB 1021 DAEEYLVPOQGFCCPDPAAGACGMHHRSSSTSSGCGDLTLGLEBEEAPSPPLAPS 1080
 QY 1071 EGAGSDVPDGLDMGAKAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPPE 1130
 DB 1081 EGAGSDVPDGLDMGAKAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPPE 1140
 QY 1131 YNQPDPVRPQPSPREGLPAPRPGATILERAUTLSPKNGVYXVDFAGAVENPEYLT 1190
 DB 1141 YNQPDPVRPQPSPREGLPAPRPGATILERAUTLSPKNGVYXVDFAGAVENPEYLT 1200
 QY 1191 POGGAPOHPPPAPSPARFNNLYYNDOPPERGAPSPFETKPTTAENEYGLDVPV 1247
 DB 1201 PREGTASPPHPSPAPSPARFNNLYYNDOPPERGAPSPFETKPTTAENEYGLDVPV 1257

RESULT 3
 ERB2_MESAU
 ID ERB2_MESAU STANDARD; PRT: 1254 AA.
 AC 06053;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 OS ERB2 OR NEU.
 GN Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 NC NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene";
 RT Gene 140:251-255(1994).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUUS.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL; D16295; BAA03801.1; -
CC HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; EGF_Kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_Kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 1 21
FT DOMAIN 22 1254
FT TRANSMEM 22 652
FT DOMAIN 653 675
FT DOMAIN 676 1254
FT DOMAIN 158 368
FT DOMAIN 472 644
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 236 244
FT DISULFID 240 252
FT DISULFID 255 264
FT DISULFID 268 295
FT DISULFID 299 311
FT DISULFID 315 331
FT DISULFID 334 338
FT DISULFID 511 520
FT DISULFID 515 528
FT DISULFID 531 540
FT DISULFID 544 560
FT DISULFID 563 576
FT DISULFID 567 584
FT DISULFID 587 596
FT DISULFID 600 623
FT DISULFID 626 634
FT DISULFID 630 642
FT MOD_RES 1139 1139
FT MOD_RES 1247 1247
FT CARBOHYD 68 68
FT CARBOHYD 125 125
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 658 658
FT VARIANT 659 659
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.6%; Score 5799.5; DB 1; Length 1254;
Best Local Similarity 85.3%; Pred. No. 9.5e-304;
Matches 1070; Conservative 64; Mismatches 112; Indels 9; Gaps 4;

QY 1 MELALCRMGILLALPPGAASVCTGTDMKRLRLPASPEHLDMLRHLYGCGVVGNTL 60
DB 1 MELAMCGMGLLALLSPGASGTGCTGTDMKRLRLPASPEHLDMLRHLYGCGVVGNTL 60

QY ELTYLPNNAISLFLQD:QEVGYVLLAHNOVROVPLORLRYRGTOLEFEDNYALAVLQNG 120
DB ELTYLPANATISFLQD:QEVGYVLLAHNOVROVPLORLRYRGTOLEFEDNYALAVLQNG 120
QY DPLNNTTPVTGASFGRLQLRLSTLEILKGVLLGRNPOLCYQDTITLWKDIFPKKNOLA 180
DB DPLNNTTPVTGASFGRLQLRLSTLEILKGVLLGRNPOLCYQDTITLWKDIFPKKNOLA 180
QY LTLIDNRSRACHPCSMCKSGRCKWESSSDQSLRTYVAGGACAKGRLPTDCHQEC 240
DB LTLIDNRSRACHPCSMCKSGRCKWESSSDQSLRTYVAGGACAKGRLPTDCHQEC 240
QY AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTESMNPGRGRTFGASCYTAAP 300
DB AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTESMNPGRGRTFGASCYTAAP 300
QY YNLTSTVGSCTIVCPILHNOETVAEDGTORCEKSPCARVYCYGLGMQYIKANSKFIGIT 360
DB YNLTSTVGSCTIVCPILHNOETVAEDGTORCEKSPCARVYCYGLGMQYIKANSKFIGIT 360
QY ELFPAGCKTIFGSLAF:PFESFGDDPASNTAPLOPDLQVETLEETGYLXISAMPDISLP 420
DB ELFPAGCKTIFGSLAF:PFESFGDDPASNTAPLOPDLQVETLEETGYLXISAMPDISLP 420
QY DLSVFGNQLVIRGRILHNGAYS:LTLOGLSIMLGRLSRLRELSGGLALIHNTHLCFHTV 480
DB DLSVFGNQLVIRGRILHNGAYS:LTLOGLSIMLGRLSRLRELSGGLALIHNTHLCFHTV 480
QY PMOQLPRNPQALHRTANRPEDECVESGIALCHQLCARHGMGPGPTQVNCGQFLRGQEC 540
DB PMOQLPRNPQALHRTANRPEDECVESGIALCHQLCARHGMGPGPTQVNCGQFLRGQEC 540
QY VEECRVLOGLPREYVNAHCLPCHPECPQNSGVTGFEADQCAVAYKDPRECVARC 600
DB VEECRVLOGLPREYVNAHCLPCHPECPQNSGVTGFEADQCAVAYKDPRECVARC 600
QY VKECRVWKGLPREYVNAHCLPCHPECPQNSGVTGFEADQCAVAYKDPRECVARC 600
DB VKECRVWKGLPREYVNAHCLPCHPECPQNSGVTGFEADQCAVAYKDPRECVARC 600
QY PSQVCKDLSMPTWKFPDEEGACQPCPINCSTHSCVDLDRKGPABORASPLTISAVAG 660
DB PSQVCKDLSMPTWKFPDEEGACQPCPINCSTHSCVDLDRKGPABORASPLTISAVAG 660
QY ILVWVVGVPFGLI:---QYIKANS--KFIGITEL--PLTSGAMPQAOQRILKEITEL 712
DB ILVWVVGVPFGLI:---QYIKANS--KFIGITEL--PLTSGAMPQAOQRILKEITEL 712
QY ILVWVVGVPFGLI:---QYIKANS--KFIGITEL--PLTSGAMPQAOQRILKEITEL 712
DB ILVWVVGVPFGLI:---QYIKANS--KFIGITEL--PLTSGAMPQAOQRILKEITEL 712
QY RYKVKLGSAFGTVYKGIWIPGENVVKIPVALIKVLRNTSPKANKELIDEAYVAVAGVSP 772
DB RYKVKLGSAFGTVYKGIWIPGENVVKIPVALIKVLRNTSPKANKELIDEAYVAVAGVSP 772
QY RYKVKLGSAFGTVYKGIWIPGENVVKIPVALIKVLRNTSPKANKELIDEAYVAVAGVSP 772
DB RYKVKLGSAFGTVYKGIWIPGENVVKIPVALIKVLRNTSPKANKELIDEAYVAVAGVSP 772
QY YVSRLLGICLTSTVQLVQLMPYGLLHVBEHRRLQSDLLMWCVQIAKMSYLEEVR 840
DB YVSRLLGICLTSTVQLVQLMPYGLLHVBEHRRLQSDLLMWCVQIAKMSYLEEVR 840
QY LVHRDLAANLVKSPNHNKIDTFLARLLDIDETRYADGAKVPIKMALESLIRRF 892
DB LVHRDLAANLVKSPNHNKIDTFLARLLDIDETRYADGAKVPIKMALESLIRRF 892
QY LVHRDLAANLVKSPNHNKIDTFLARLLDIDETRYADGAKVPIKMALESLIRRF 892
DB LVHRDLAANLVKSPNHNKIDTFLARLLDIDETRYADGAKVPIKMALESLIRRF 892
QY HOSDVSWSGVYTWELMTGAKPYDGI:PAEIPDLLEKGERLPQPPICITDYYMIMVCM 952
DB HOSDVSWSGVYTWELMTGAKPYDGI:PAEIPDLLEKGERLPQPPICITDYYMIMVCM 952
QY HOSDVSWSGVYTWELMTGAKPYDGI:PAEIPDLLEKGERLPQPPICITDYYMIMVCM 952
DB HOSDVSWSGVYTWELMTGAKPYDGI:PAEIPDLLEKGERLPQPPICITDYYMIMVCM 952
QY ISECRPRPRELVSFESMARDDPQRFVYI:QEDIGPSSPLDSTYRSLLEDDMDGLVDA 1012
DB ISECRPRPRELVSFESMARDDPQRFVYI:QEDIGPSSPLDSTYRSLLEDDMDGLVDA 1012
QY ISECRPRPRELVSFESMARDDPQRFVYI:QEDIGPSSPLDSTYRSLLEDDMDGLVDA 1012
DB ISECRPRPRELVSFESMARDDPQRFVYI:QEDIGPSSPLDSTYRSLLEDDMDGLVDA 1012
QY BEYLVPOGQFFPCPDAPAGGVVHRRSSSTRSGGDLTLGLPSESEEA:PRSLAPSEG 1072
DB BEYLVPOGQFFPCPDAPAGGVVHRRSSSTRSGGDLTLGLPSESEEA:PRSLAPSEG 1072
QY AGSDVFPDDGLMGAKAGLQSLPTHDPSPLQYISDPTVPLPSEFTDGYVADLTGSPQGEYV 1132
DB AGSDVFPDDGLMGAKAGLQSLPTHDPSPLQYISDPTVPLPSEFTDGYVADLTGSPQGEYV 1132
QY AGSDVFPDDGLMGAKAGLQSLPTHDPSPLQYISDPTVPLPSEFTDGYVADLTGSPQGEYV 1132
DB AGSDVFPDDGLMGAKAGLQSLPTHDPSPLQYISDPTVPLPSEFTDGYVADLTGSPQGEYV 1132

```

Cy 1133 NOPDVRPQPSPREGLPAPARPGATLERAKTLSPKNGVYKDVFAFGAVENPEYLTPO 1192
Db 1141 NQPEVRPQPLTEGGLPVPVRPGATLERKTLSPKNGVYKDVFFGGAIVENPEYLVPR 1200
Cy 1193 GGAAPQPHPPAPSPADNLYYNDQDPPERGAPSPFTFKGPTTANPEYIGLDVPIV 1247
Db 1201 GGSASQPH-PPALCPAFDNLNLYWDQDPSERGSPPNTEGPTAENPEYIGLDVPIV 1254

RESULT 4
EGFR_HUMAN STANDARD; PRT: 1210 AA.
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC Q00688; Q9BZ52; Q9H2C9; Q9H2X1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilkies J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maibach N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilkies J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=1161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Scheel Sinclair C., Pearisall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maibach N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Scheel C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

```

```

RA Maibach N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger M., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Stamen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Walley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:1125-1132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-386(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Howk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 268:10667-10671(1993).

```

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renout D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-glycosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p66/
 CC truncated isoform/TEGR, 3/p110 and 4, are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00588; CA25240.1; -
 DR EMBL: U95089; AAB53063.1; -
 DR EMBL: U48722; AAC50802.1; -
 DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -
 Query Match 45.2%; Score 3064; DB 1; Length 1210;
 Best local similarity 49.0%; Pred No. 4, 6e-157;
 Matches 619; Conservative 177; Mismatches 357; Indels 110; Gaps 24;

QY ASLSFLDIOEVQGVYVLIANNOVQVPLQRLIRVSTGTOFEDNVVALAVDNGPPLNNTTP 128
 DB YDSFLKTIQEVAGVYVLIANTVERIPLENLQIRNMYEENSVALAVLSND----- 126
 QY 129 VTGASPGGLBELQRLSTLTKKGVLIQNNPOLCYDDTLMKDIFKKNQALTLIDTR 188
 DB 127 ---ANKTKLELPMRNLQETLHVAVFSNNPALCNVESIQMRDIVSSDLNMSMDFQNH 183
 QY 189 SRACHPCSPMKCKSRGCGESSBQCSGLTRTVCAAGCA-CKGGLPFDCCHEQCAACTGP 247
 DB 184 LGSQKDPSCPCNPGSCMGAGEHCQYLTKICAQQSGCGRCGSPBDCCHNQCAACTGP 243
 QY 248 KHSDDCALHFHNSGICELHCPALVTYNTDTFESMNPBGRYTFGASCVYACPPYVLS 307
 DB 244 RESDCLVCRKFRDEATCKDTPMLNPTTYQMVDVNPBGRKTSFGATCVKCKPRNVVTD 303
 QY 308 VSGCTLVCPHNOEVAADGTQCEKSCSPKARVYGVGLQMOYIKANSKRTIGTELE-FAG 366
 DB 304 HGSVCRAAGADSTEM-EBDGVKCKCKCEGPKRVKNGIGIGEFK-DLSINATNIKHFN 361
 QY 367 CKKIFGSLAFLPSPFDGDPASNTAPLOPEQLOVFTLEETITGLYISAMPDLSPLVQ 426
 DB 362 CTSISGDLHLIPVAFRGDSFTTHPPDPOELDIKTVKEITGFLLIQAMPKRTDLHAF 421
 QY 427 NLQYIRGRIILHNGAYSLTQGLGISWLGRSRLBAGSGLAIHHNTHLCFHTVPWDL 486
 DB 422 NLBIIRGRTKQHQFSLAVVSLNITSLSGRSLKEISDGDVIVISGNKNLCYANTINWKLF 481
 QY 487 RNPHOALHTANRPEBCEVGEGLAQQLCARHCWGPGPTQCVNCSQFLRQCEBCECV 546
 DB 482 GTGQCKTKIISNGENSCATGQVCHALSPGCGPBERDVCNWSRGRECVDKWL 541
 QY 547 LQGLPREYVNAHCLPCHPECOPONGSVTCFPEADQCAAHYKDPFCVARCPSGVK 606
 DB 542 LGEPRFEVENSICQICHECPLQANNTICTGRPDNQCAGHYIDGRCVYCTCAGVWG 601
 QY 607 DLSYMIKWFPEBEGCQCPINCHTSCVDLDDKCGPRAQASPLTSVSAVG--ILL 663
 DB 602 ENNTL-VMKYADAGHVCHLCHNCTYGCGLPGLBCPTNGPKIP--SIATGVAGLALLL 658
 QY 664 VVVLGVVFGILIOY---KANSKFTIGTEL--PLTPSGAMPQACMRILKETELKRVYL 718
 DB 659 VVALGIGLMRRRHRIYKRTLRLLQERLVEPLTPSGAPQALRLKETEFKRVYL 718
 QY 719 GSGAFGTVYKGIWIPDENVKIPIVAKVIRENTSPKANKEILDEAYVAGVSPVSRLL 778
 DB 719 GSGAFGTVYKGIWIPGEKVKIPIVAKELREATSPKANKEILDEAYVAVASVDNPHVCEL 778
 QY 779 GICLTSTVQLVQLMPYGLDHYVENGRGLSGODLLMWQCIAGMSYLEDVRLVHDL 838
 DB 779 GICLTSTVQLVQLMPYGLDHYVEHKNQISQYLLMWQCIAGMNYLEDRLVHDL 838
 QY 839 AARNVLVKSNNHVKITDFGLARLIDIDETEHADGKVPKIMMALESILRRRFTHQSDW 898
 DB 839 AARNVLVKTQGVKITTDFGLARLIDAEKEVHAEGCKVPIKMMALFSLIHLIYHQSDW 898
 QY 899 SYGVYTWELMTFGAIPYDGIIPAREPLDLKEGRIPPPCTIDVYIMIVKGMIDSECR 958
 DB 899 SYGVYTWELMTFGSKPYDGIIPASEISLLEKGERLPDPCTIDVYIMIVKGMIDADR 958
 QY 959 PRFRELVSFSSMADDPORFVYIQ-NEDLGPASPLDSFFYSLLDDMGDLVAEEV 1017
 DB 959 PKFRELIIEFSQADDPQRYLVYIQDENHMLPSPDLSFFYALMDDEMDVDVADDEYL 1018
 QY 1018 PQGGFPCDPAPGAGVYVHHRRSSSTRSGGDLTLGLPSEEBAPSPPLAPSEAGSDV 1077
 DB 1019 PQGGF-----SSPSTRPLSSLSATSN- 1043
 QY 1078 PDGDDGMAKAGLQSLPTHDSPLQRYSEDPVLPBRT--DGVAPLTCSPQREYVNP 1135
 DB 1044 -NSTVACTIDRNGLOSCPIKEDSFLQRYSSDPTGALTDSIDDTFL-----PVPEYINQ- 1095

QY 1136 DVPQPSPPREGPLPAARPAQATLERAKTLPKQNGVWQVAFGAVENPEYL-TPQCG 1194
 DB 1096 SVPRPAGSVQNPVYHQNLP-----APSRDPHYQD--PHSTAYQNPETLNVQ-- 1143
 QY 1195 AARPPHPAPSPAFNDLYWDQ-----DP-----PRGAPPSFTFKTPTAENP 1238
 DB 1144 -----PTCVNSTFDSPPAHMAQGHQISLNDPDYQDFFPEAKRNQIPFGS-TAENA 1195
 QY 1239 EYL 1241
 DB 1196 EYL 1198
 RESULT 5
 EGF_MOUSE
 ID EGF_MOUSE STANDARD; PRT; 1210 AA.
 AC 001279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGF.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yaron A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 RT (bek/KGFR) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 RT in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 RT receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Taxt I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 RT binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Eslinger D.P., Serrero G.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X79987; CA55587.1; -
 DR EMBL; U03425; AAA1789.1; -
 DR EMBL; X59698; CA44219.1; -
 DR EMBL; L06864; AAA5302.1; -
 DR EMBL; Z12608; CAAT8249.1; -
 DR HSSP; P11362; 1FGK
 DR MGD; MGI:95294; Egfr.
 DR InterPro; IPR000494; EGF_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; pkinase; 1.
 DR Pfam; PF01030; Recep_L-like; 1.
 DR Pfam; PD000001; Euk_pkinase; 1.
 DR ProDom; SM00261; Fu; 3.
 DR SMART; SM00219; TYKc; 1.
 DR SMART; SM00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 DR Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT NP_BIND 714 981
 FT BINDING 720 728
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 336
 FT DISULFID 329 333
 FT DISULFID 506 515
 FT DISULFID 510 523
 FT DISULFID 526 535
 FT DISULFID 539 555
 FT DISULFID 558 571
 FT DISULFID 562 579
 FT DISULFID 582 591
 FT DISULFID 595 617
 FT DISULFID 620 628
 FT DISULFID 624 636
 FT MOD_RES 660 680
 FT MOD_RES 1092 1092
 FT MOD_RES 1092 1092
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT	MOD_RES	1110	1110	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1172	1172	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1197	1197	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	175	175	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	413	413	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	603	603	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CONFLICT	539	539	C -> S (IN REF. 2).	
FT	CONFLICT	591	591	C -> W (IN REF. 5).	
FT	CONFLICT	991	991	L -> F (IN REF. 4).	
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6).	
SO	SEQUENCE	1210	134853	MM; 690E20D46DFD2F5	CRC64;
Query Match					
Best Local Similarity		48.84;	Score 3033;	DB 1;	Length 1210;
Matches 621;		Conservative 170;	Mismatches 363;	Indels 118;	Gaps 26;
QY	11	LLALLPAGAA--STQVCTGTDMKLRPASPEHLDMLHLYQGGQVQGNLELTLYPTN	68		
DB	14	LLTALCAAGALEKKKVCOSTSNRLTQGTFFEDHFLSLQRMNNEVGLNLEITVQGN	73		
QY	69	ASLSFLQDIOEVQVYLIAANQVQVPLQRLVSTQGLFEDNYALAVLNDGPDPLNTP	128		
DB	74	YDLSEFLKTIQEVAGVYVIALNTERIERLENQIRNALYENTVALALISN-----	124		
QY	129	VTGASPGALRELOLRSTELIKGVLIOBPOLCYODTILMKDI----FHKNNQALTLI	184		
DB	125	-YGNRTGLRELPMKNGOELLIGAVRFSNNPILCMNDITQMDIYQVNFMSMDL----	180		
QY	185	DTNRSRACHPCSPCKSCRGWSSSDQSLRTVCAGCA-RCKGRLPTDCCEQCAAG	243		
DB	181	-QSHPSRCPCDSCPCWSSCGGEBENCOKTKIICAOQCSHRCSGRSPSCCHQCAAG	239		
QY	244	CTGPKHSDCLACHFNHNSGICELCPALVTYNDTPESMNPNEGRTTGASCVTAAPYNY	303		
DB	240	CTGPRESDCLVCCQKFOEATCQTCQCPPLMYNTTYQMVNNEGKYSQATCVKCPYNY	299		
QY	304	LSTDVSGCTLVCPHLNHOVTAEDGTQCEKSCPKARVCYGLQWQYIKANSKFIGITELE	363		
DB	300	VYTDHGSQVACGPDYEV--EEDGIRKKCKCGPCCKVCNGIGIEFK-DLISINATWIK	357		
QY	364	-FAGCKKIFGSLAFLEPSFDGPDASNTAPLOPELOVFTLEITGYLYISAMPDLPDL	422		
DB	358	HFKYCTAISGDLHLPLAFKGBDFTRTPDRELEILKTVKEITGFLLIQAMPDNTDL	417		
QY	423	SVFQNLQVINGRILHNGAYSLTQGLISWLGISRLREIGSGALIHNTLCEVHTVPM	482		
DB	418	HAFENLEIIRGRTKQHQFSLAVAGLINTSLGIRSLKEISDDDVIIISGRNLQYANTINW	477		
QY	483	DQCFRNPHOALHNTANPREDECVGEGLAGHOLCARHGCPPTCCVACSCQPLRQOEVE	542		
DB	478	KKLRTGPNQKTKIMNRAEKDCANVHNCVPLCSSGCGPPEPCVSCQVNSRRECEVE	537		
QY	543	ECRVLAGLPREVYNAHRLCPHECPQNGSVTCFGEADQCVACAHYKDPFCVACPS	602		
DB	538	KCNILEGEPRFEVENSECICQHECPLOAMNITCTGRGDNDICCAHVIDGHCVKTCFA	597		
QY	603	GVRPDLSTYMPIMKRPDEGACQCPINCNSCVDDLDKCPAEQASPLTSLVAVVUIL	662		
DB	598	GIMENNTL-VWKYADANNVCHCANCTYGCAGPQLQCEVWBSGPKIPSLAIGVGL	656		
QY	663	L--VVVLGVFGILLQYI--KANSKFIGITEL--PLTPSGAMPNQAQMRILKETELRK	714		
DB	657	LFIYVVALGIGLFWRBRHIVRKRTLRLOEBELVPLTPSGEAPNQAHLRIKETEKX	716		
QY	715	VKVLGSAFGVYVKKGIIPDGENVKIPVAIKVLRNTSPKANKELIDDAVYVAGVGSYV	774		
DB	717	IKVLGSAFGVYVKKGIIPDGENVKIPVAIKVLRNTSPKANKELIDDAVYVAGVGSYV	776		
QY	775	SELLGICLSTVQVLTQMPYGCILDHVRENRLQSGODILNMCQIAGMSLEDEVRIV	834		
DB	777	CLULICLSTVQVLTQMPYGCILDHVRENRLQSGODILNMCQIAGMSLEDEVRIV	836		
QY	835	HRDLAARVAVLKTQHVXITDFGLAKLGAEXEYHAGGKVPDKMALESILHRTYHQ	894		
DB	837	HRDLAARVAVLKTQHVXITDFGLAKLGAEXEYHAGGKVPDKMALESILHRTYHQ	896		
QY	895	SVWASGYVTWELMTFGAKPYDGIIPAEIFDLLEKGRLLPQPICTIDVYMTVKKMMID	954		
DB	897	SVWASGYVTWELMTFGAKPYDGIIPAEIFDLLEKGRLLPQPICTIDVYMTVKKMMID	956		
QY	955	SECRPREFELVSEFSRMAPDQRFVYIQ-NEDGAPASPLDSTFYRSLLEDODGVLDAE	101		
DB	957	ADSRKPFRELLIFSRYARQDQRLVIGQSERHNLPSPTSNFYRLMDEDMEDVVDAD	101		
QY	1014	EYLVVQGGFFCFDPAPGAGMHHRRSSSTRSGGDLTIGLEPSEEARPSFLASBEA	107		
DB	1017	EYLVVQGGFFCFDPAPGAGMHHRRSSSTRSGGDLTIGLEPSEEARPSFLASBEA	107		
QY	1074	GSDVDCDGLGMAAKGQSLUPTHDSPDQRYSEDPVPLPSET--DGVAAPLTCPOPEY	113		
DB	1043	TSN-----NSTVACINRNSGCVKEDATLQYSSDPFGATVEDNIDAFI-----PVPEY	109		
QY	1132	VNQPDPVFPQPSPPBGPPLAPAPGATLEAKTILSPCKNGVVKDVARFGAVNPEYL-T	1191		
DB	1093	VNQ-SVPRKPAQSVQNPVYHQPLHP-----APGRDLHYQN--PHSNAQNPPEYLT	114		
QY	1191	PGGAAPQHPHPAPSPAPDLYWDC-----DP-----PERGAPSTFGCTPT	123		
DB	1142	AQ-----PTLSSQSPNPAIMQKSHQMSLNPQVQDFKTKRPNQIFRG-FT	1191		
QY	1235	AENPEYGLDVP	1246		
DB	1192	AENAEYLRVAP	1203		
RESULT 6					
ERBA_HUMAN	STANDARD;	PRT;	1308	AA.	
ID	Q15303;				
AC	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Receptor protein-tyrosine kinase erbB-4 precursor (FC 2.7.1.112)				
GN	ERBB4 (tyrosine kinase-type cell surface receptor ERB4)				
OS	ERBB4 OR HER4				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
NCBI	Taxid:9606;				
RP	SEQUENCE FROM N.A. (ISOFORM JM-A).				
RC	TISSUE=breast carcinoma				
RX	MEDLINE:93189574; PubMed:8393326;				
RA	Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,				
RA	Foy L., Neubauer M.G., Shoyab M.;				
RT	"Ligand-specific activation of HER4/p10erbB4, a fourth member of the				
RT	epidermal growth factor receptor family."				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).				
RP	SEQUENCE FROM N.A. (ISOFORM JM-B).				
RC	TISSUE=petal brain;				
RX	MEDLINE:97476287; PubMed:9334263;				
RA	Eleinius K., Cortes G., Paul S., Choi C.J., Rio C., Plozman G.D.,				
RA	Klaesbrun M.;				
RT	"A novel juxtamembrane domain isoform of HER4/erbB4, isoform-specific				
RT	tissue distribution and differential processing in response to				
RT	phorbol ester."				
J. Biol. Chem.	272:26761-26768(1997).				

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NR-
 CC 2. NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTX. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L07868; AAB59446.1; -
 DR HSSP: P11362; 1FGK.
 DR Genew: HGNC:3432; ERBB4.
 DR MIM: 600543; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; Fur_4.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT NP_BIND 724 732
 FT BINDING 751 751
 FT ACT_SITE 843 843
 FT DISULFID 189 197
 FT DISULFID 193 205
 FT DISULFID 213 221
 FT DISULFID 217 229
 FT DISULFID 230 238
 FT DISULFID 234 246
 FT DISULFID 249 258
 FT DISULFID 262 289
 FT DISULFID 293 304
 BY SIMILARITY.
 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 CYS-RICH.
 PROTEIN KINASE.
 ATP (BY SIMILARITY).
 ATP (BY SIMILARITY).
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1256 1256 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NPTSHDITPPYPMGHSTLPQHA -> IGSSIEDICGLMD
 (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE8095D8B761 CRC64;
 Query Match 43.0%; Score 2915.5; DB 1; Length 1308;
 Best Local Similarity 44.7%; Pred. No. 4,6e-149;
 Matches 603; Conservative 183; Mismatches 377; Indels 185; Gaps 30;
 QY 9 KGLLALLPQGA-----STVCTGDMKLRLPASBETHLMDLRLHYQCCVQVQNTLETY 64
 DB 8 WWSVLVAAGTVQPSDSQSVACAGTENLSSLSDEQYRALRYENCEVVMGNLEITS 67
 QY 65 LPTNASTLPIDICQVGVVIAHQVQVLORLRIYRGQLDEEDNVALAVLDNGPBLN 124
 DB 68 IHNNDLSLTRVREVTYVVALNQFYFLLENRIIRIGKYLEDRALALFLNYRKDG 127
 QY 125 NTPYTGASPGGLRELOSLTEILKGVLIQRNPOLCYOFTILMKDIFHAKNOLATLI 184
 DB 128 NF-----GLQELGKMLTEILNGVVYDQKKFLCYADTIHMQDIVNPMPSNLTIV 178
 QY 185 DNNRSPACHPCSPMKSGRCWGESSEDDQSLTRVYACGC-ARCKGPLPDDCCHEQCAAG 243
 DB 179 STNGSSGGGRCHKSGCTG-RCWGPLENNHQUTLRTVCAEQCGRCYGVASDCCHEBCAGG 237
 QY 244 CTGPRGSDCLAFHNHSGICELCPALVTYNTQTFESMPREGRYFGASCVTACPYNY 303
 DB 238 GSGPDDTCFACNMNNDGACVTCQPFVFNPTTFQLEHNFNAKTYGAFCYKCKCHNF 297
 QY 304 LSTDVSGCTLYCPLHNOVYTAEDGTORCEKSKPCARVYCGLGNQYIKANSKFITTELE 363
 DB 296 V-VDSSTCVRACPSPSKREV-EENGIRKMKPCTDICPRACDGIIGTGLMSAQTVDSNIDK 355
 QY 364 FAGCKRFGSLAFIPESPDPDPAANTAPLOEPOQVETLEITGVYISAMPDLSLPS 423
 DB 356 FINCRIKINGNLIPLVTGTHGDPYVAIEADBEKLVNRTYREITGFANISWPPNMTDFS 415
 QY 424 VFQNIQVIRGILNNGAVSLTQGLGIWGLRLRLSGGLALIHNNTHCFVHTVPMW 483
 DB 416 VFSNLTITGVHLYSGSLILTKQOGLTSLOFGLKEISAGNIYITNSNLCGYHTINMT 475
 QY 484 QLEFRPHQALHTNRPDEVCVGGGLACHQLCARHGWGGRPPQCVNCSFPLGQCEVBE 543
 DB 476 TLFTSTINRIYIRNRAENCTAGWVCNHLCSDDGWGSPQCCSRFSRGRICIES 535
 QY 544 CRVLQGLPREVYVNAHCLPCHPECP-ONGSVTCFGEADQCVACAAYKPPPCVARCPS 602


```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 ATP (BY SIMILARITY).
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

```

Query Match 42.7%; Score 2896; DB 1; Length 1308;
 Best Local Similarity 44.6%; Preg. No. 5,1e-148;
 Matches 601; Conservative 190; Mismatches 386; Indels 172; Gaps 30;

```

QY 1 MEIA-ALCRWGLL--ALLPQASTQVCTGDMKLRLPASPTHLMRLHYGCGQVQ 57
DB 1 MKLATGLWVGSLVAARTVQPSASQSCAGTENKLSLSLEQYRALRKYIYENCEVM 60
QY 58 GNLELTLYPTNASISFLDIOGVGVLIANNQVQVLRQLRVKGTQLFEDNYALAVL 117
DB 61 GNLEITISIEHNDLSFLRSIREVTGYVALNQPFIYPLENLRIRIKYLEDYALAF 120
QY 118 DNGDEPLNNTTPTVTCASPGSLRELOSLTEILKGGVLIQBNPQLCYDITLMKDIHKFN 177
DB 121 LNYRKDGNF-----GLQELGAKNTEILNGVYVDQKFLCYADTTHWQDIYANPW 171
QY 178 QIALTLIDTNRSRACHPCSPMKCGSRGSESSDQSLTRTVCAAGC-ARCKGLPTDCC 236
DB 172 PSNMTLVSTIGSSGGRCHKSGCTG-RCWGPTEHHCQTLTRTVCAEQDQGRGYPYVSDC 230
QY 237 HECCAAGCTGPKHSCCLAHFNHSGIGELHCPALVLYNTDTFESMNPBGRITTFGASCV 236
DB 231 HRECAAGCSGPKDITCFACWMPNFGACVTCQPPQTFYVNTPTFLFNFAKXTYGFVFCV 290
QY 297 TACPNYVLTSDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVAVCYGLMQYIKANSKF 356
DB 291 KKCPHNFV-VDSSSCVRCACPSKMEV--EENGITKMKCRCTDICPAACGIGTGSIMAGQV 348

```

```

QY 357 IGTELEFAGCKKIFGSLAFLEPESFDGDPASNTARLOPEQLQVETLEETGYLYISAMP 416
DB 349 DSNINIKFNTCKINKNLIFLTGTHGDPYNAIDAIDEKLVETFEVITGFLNIQTWF 408
QY 417 DSLPLSVFQNLQVIRGRILNHAVALTLQGLISWLGSLRSLGSLALIHNTHLCF 476
DB 409 PNMTRDSVSNLVITIGRVLYSGLSLILKQGGISLQFOSLKEISACNIYITDSNICY 468
QY 477 VHVVPDQLEFRPHQALHTARPEDECVGSEGLACHQICARHCGCPPTQCVNCSQFLR 536
DB 469 YHTINMTLTFSTVNRIVIRDRRAENCTAEGMVCNHLCSNDGCMGPGPDQCLSCRRFSR 528
QY 537 GQECVEECVTLQGLPREYVNAHRCPCPCPECP-ONGSVTCFPGPADQCVACAYKDPDF 595
DB 529 GKICIESCNLYGGEFREFPENGSIYCECCSQCKEMEDGLITLHGPPDNTCASHFKDQPN 588
QY 596 CVARCPGVKPLSTWPIWKFPDEEGACQPCPINCTHSCVDLDDKC-----PA 644
DB 589 CVEKCPDVLQGANSE--IFKYADQDRCHPCHPNCTQCGNPGTSHDCIYPMGTGHTLPQ 646
QY 645 EDRASPLTIVSAVY-GILLVVLGVGVGILLIYIKANS-----KFIGITEL--PLT 693
DB 647 HAR-TPL--IAAGVIGSLVITVMTALTAV--VYRRSIIKKRALKALRFLF-TLVEPLT 699
QY 694 PSQAMENQAKRILKETEELRKYKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLENTSP 753
DB 700 PSQTAPOAQNLILKETELKRYKVLGSGAFGTVYKGIWIPGEVTKIPVAIKINLETTGP 759
QY 754 KANKELDEAVYMAVGSPYYSRLIGTLTSTVOLVQIMYGCILDHVRENRLCSOD 813
DB 760 KANVEMEDALIMASVDHPHVRLLGVLSPTLQVQIMPHGCLLEVHHEKXNISQOL 819
QY 814 LNMGCQIAKAGSYLEDVRLVHRDLAARNVLYKSPNHKIDFGIARLDDDETEYADG 873
DB 820 LNMVCQIAKAGMYLEERLVRHDLAARNVLYKSPNHKIDFGIARLDDDETEYADG 879
QY 874 GVPIPKMALESILRRFTTHOSDVASVYTWELMTFGAKPYDGIIPAREIPDLLEKGERL 933
DB 880 GMPPIKMALECIHRYKFTTHOSDVASVYTWELMTFGAKPYDGIIPAREIPDLLEKGERL 939
QY 934 POPPICTIDVYIMYKCMWIDSECRPRELVSFESMARPQRFVYIIONED-LGPASPL 992
DB 940 POPPICTIDVYIMYKCMWIDSDSPKKELEAERSRARPQRFVYIIONED-LGPASPL 999
QY 993 DSTFYSLLIEDDMGLVDAAEYLVPOQGFPCPD-----AP 1029
DB 1000 DSKFQNLDDDEDDMDMADELYLP-QAFNIPPIYTSRTRIDNSNSELGHSPPAYTP 1058
QY 1030 GAGGMYHHRHSSSTRSGGDLITGLBSEDEARSLAPSEAGSVFPDGLGMAKAG 1089
DB 1059 MSGSQFYQDGFATQCG--MPMPYATATSTIEAFAVA--QGATAEMFDDSCNGTLRK 1113
QY 1090 LQSLPTHPSPFORSEDPVPLPS-----ETDGVAPLPSCPOEYVNOPDVRQPP 1142
DB 1114 PVPVHVQDSSDTQKYSADPTFAFERNPRAPLDEBEGIMTGMHDKPKKEYLYNPVE----- 1167
QY 1143 SPREGPLPAPAPAGATLERAKTISPGKNGVYKDVFAFGAVENPEYLTPOGGAAPQHP 1202
DB 1168 ---ENPFVSRR-----XNGDLC-----ALDNPETHASASSG-----PP 1196
QY 1203 PA-----PSAPFDULYWDQPPERGA--PP 1226
DB 1197 KAEDEYVNEPLVNTFTNALGNAEYMKNSLISPEKAKKAFNDPDMWNHSLPBRSTLQHP 1256
QY 1227 STFKGTPT-----AENPEYL 1241
DB 1257 DYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

RESULT 8
 XMRK_XIPMA
 ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;

DT 01-JAN-1990 (Ref. 13, Created)
 DT 16-OCT-2001 (Ref. 40, Last sequence update)
 DT 15-JUN-2002 (Ref. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malischek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.W., Schartl M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1998).
 RN [2]
 RP REVISION TO 515.
 RA Schartl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X16891; CAA34770.2; -.
 DR PIR; S06142; S06142.
 DR HSSP; P1362; 1FGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PFO0069; kinase; 1.
 DR Pfam; PFO0757; Furin-like; 1.
 DR Pfam; PFO1030; Recep_L_domain; 2.
 DR PRINTS; PRO0109; TYRKINASE.
 DR ProDom; PDO00001; Euk_pkinase; 1.
 DR SMART; SMO0261; FU; 5.
 DR SMART; SMO0220; S_TKc; 1.
 DR SMART; SMO0219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 643 665 POTENTIAL.
 FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 710 977 PROTEIN_KINASE.
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 ATP (BY SIMILARITY).
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 254 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 123934 MW; 4793E0749DC1D55A CRC64;
 Query Match 38.6%; Score 2612.5; DB 1; Length 1167;
 Best Local Similarity 44.3%; Pred. No. 7.2e-133;
 Matches 561; Conservative 167; Mismatches 394; Indels 145; Gaps 30;
 QY 4 AALCRWGLLALLPFGAAS-----OVCTGTMKRLPASPEHMLDMLRGOGGVQVGN 59
 DB 8 AALLQ--LLLVISIRCCSTDPDRVCQGTSNQM--LNNHLYMKMKMSGGCNVLEN 62
 QY 60 LELTYLPYNASLSPLQDIOEVGYVLLAHNOVROYPLQRLRIVERGTOLFEDNYALAVLDN 119
 DB 63 LEITYTOENODLSFLQSIQEVGYVLLAHNEVSTPLVNLRLIRIGONLYEGNFFLLWMSN 122
 QY 120 GDPLNNTPTVGASPGRLRELQSLSTELIKGVLLIQNPOLCYQDTILMKDIFHKNNQL 179
 DB 123 YK--NPSSP--DIVYVGLKQQLQSLVLEISGVKVSHPNPLVCVETINMWDDVDKTSNP 179
 QY 180 ALTLIDNRSRACHPCSPCKMSCKRSMGSSSEDCSLRTVCAGGC--RRCKGPLPTDCHE 238
 DB 180 TNNLIPFAFERQCKQCHGCVNGSOMAPGPHQCKFKLLCAEQNRRCRGPRIIDCNE 239
 QY 223 QCAAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDFESNPNEGRTYFGASCUTA 298
 DB 240 HCAGGCTGPRATDCLACRFDDDCDTGCPPKXIYIVSHQVVDNENIKYTFGAACVKE 299
 QY 299 CPYNYLSTDVSCITLVCPILHNOETAEGDTGRCKSGKPCARVYGVGCMGYIANSKFIG 358
 DB 300 CPNSNVVTE--GACVRSAGAGMLEVD--ENGKRSCKPCGVCPRKCDGIGISL--SNTIAVN 356
 QY 359 ITEL--EPGCKKIKFGSLAFLPESFDGDPASNTAPLQBOQVEETLEITGYLISAMPD 417
 DB 357 STNIRSFENCTKINGDILLNRNSPEEGPHYKIGMDPEHMLNLTVEIIGYLVIMMPE 416
 QY 418 SLDPYVFNQLQVIRGRILNNGAAS--LTLOGLISWLGSLASLELBSGLALHNHNLCE 476
 DB 417 NMTLSLSPONLEILRGTTSRGGSFVVAVVRHLQWGLSLKEVAGAVILNLTQLRY 476
 QY 477 VHTVPMDQLFENPHQALLHTANRPEDECVEGLACHQLCARGCMWGGPFOCVNGSQPLR 536
 DB 477 ANTIMRRRLFRSEDSQISEYDART-----ENQTNNGESBDCMGPBPIMCVSCLHVR 529
 QY 537 GQEVVEECRVLQGLPREYVNAKCLPHEPCQPOQNSVTCFGEADQCVACAHYKDPFC 596
 DB 530 GGRCVASCNLLQGEPRFAQVDRGVCHOECLVQDTLSLTICYGPGPANCKSAHFQCGPQC 589
 QY 597 VARGSPGVKPDLSYMPIMKFPDEGACQPCPIINCTHSCVDLDDKGPAGQASPLTISVS 656

Db 590 ITRCPHGLIGDQDTLTKYKADRMGOCOPCHONCTQGSGPGLSCGRD-IVSHSSIAVG 647

Qy 657 AVVGIILVVVLGVFGILLIQTAKNSK-----FIGITEL--PLTPSGAMPQAOMRLIKE 709

Db 648 LVSGLLITIVALLIYVLLRRRIKERKTIRCLLGKEKLVEPLTPSGAPPAQALRLIKE 707

Qy 710 TELRKXVLSGAFAFVNYGMIIPGENVKIPVAIKVLRNTSPANKELLIDAYYAVAG 769

Db 708 TEKKDRVLSGAFGVYKGLMNPGENIRIPVAIKVLRNTSPANKELLIDAYYAVASV 767

Qy 770 GSPVYSRLIGLICLTSTVOYLTLMPFYGLLDHVRENRGLSODPLMWCMQIAAGMYLE 829

Db 768 DHPHVCRLLIGLCITSANVOLVTOLMFYGCILDVYQHQRICGGMLNMCVIAKGMYYLE 827

Qy 830 DYRLVHRDLAANNVLKSPNNHYKITDFGLARLLIDETEHYADGKVPDKMALESILR 889

Db 828 EHLVHRDLAANNVLLKNPNHYKITDFGLSKLLTADKEKEYAQDGKVPDKMALESILR 887

Qy 890 RPTHGSVDWVSYGATWELMTFGAKPYDGPAREIPDLLEXGERLPPOPICIDVYMIMWK 949

Db 888 YTHGSDWVSYGATWELMTFSKRKYDGPAREIPDLLEXGERLPPOPICITEYMITLK 947

Qy 950 CWMIDSECRPRELVEFSERMAARDPORFVVIYONEDLGPASPDLSTEVRSILLEDDWDGL 1009

Db 948 CMMDIPSSPRRELVEFSGMARDPSTRYLTVG---NLPELSDRLPFRSLSSDD--DV 1002

Qy 1010 VAAEYVLPQCGFFPPDPAPAGAGCVHHRHSSSTRSGGDLTLGLEPSEBEAPRSLAP 1069

Db 1003 VDADYLLPYKRI-----NRQOS-----EPCLP 1025

Qy 1070 SEGAGSDVFQDGLMGAAKGQLSFTHDSPLOLYSEDPV-PLPSETDGYVAPLTCSPQ 1128

Db 1026 PLGH-----PVRENSTILANISDPTQNALKRDLGH----- 1056

Qy 1129 PEYVNOPIVRPOP-----PSPR-----GPLP-AARPACATLERAKTLPFGKNGVWD 1175

Db 1057 -EYNQPSESTSRSLDIYNNPYEDLTDGWGPVLSQSQAETNFSPREYINTQNLSI--- 1112

Qy 1176 VFAPGAVENPEEYLLPGGAAPQPHPPAFSPADNIYYMDQDPPEEGAPPSPFKGTPTA 1235

Db 1113 PLYVSGSGDDDY---QAG-----YDAF-----LPOTGALTGNMFLPAA 1150

Qy 1236 ENPEYLG 1242

Db 1151 ENLEYLG 1157

RESULT 9
ERR3_HUMAN STANDARD; PRT; 1342 AA.
ID ERR3_HUMAN P21860;
DC 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112) (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=90083234; PubMed=2687875;
RA Kinas M.H., Issing W., Mikl T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9153-9157(1989).
RP SEQUENCE FROM N.A.
TX MEDLINE=90311312; PubMed=2164210;

PA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L., Todaro G.J., Shoyab M.;
RA Molecular cloning and expression of an additional epidermal growth factor receptor-related gene";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT c-erbB3 gene encodes secreted as well as transmembrane receptor tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A SHORT SECRETED FORM. EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annou or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -;
DR EMBL; M34309; AAA35979.1; -;
DR EMBL; S61953; AAB26935.1; -;
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR GeneW; HGNC:3431; ERBB3.
DR MiMi; 190151; -;
DR InterPro; IPRO00494; EGFR_L domain.
DR InterPro; IPRO00719; Etk_Pkinase.
DR InterPro; IPRO02174; Furin-like.
DR InterPro; IPRO01245; Tyr_Pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Etk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN KINASE_TIR; FALSE NEG.
DR TransMembran; Glycoprotein; Multigene family; Receptor; Signal;
KW Alternative splicing; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KV Alternative splicing.
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 20..643
FT TRANSMEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834
FT DISTUFD 186..194

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 566 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VASPPIC 141 183 EILSGVYIEKNDKCHMDTIDMDRIVDRRAEIVKONGR
SC -> GGFPMFSGSLTPOPADWYLLDDDDRLTLSSASK
VPVTLAAV (IN SHORT ISOFORM).
FT VASPPIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA: 148097 MM; 7201E7F6CA374ED CRC64;

```

Query Match 35.3%; Score 2390.5; DB 1; Length 1342;
 Best Local Similarity 40.2%; Pred. No. 6.8e-121;
 Matches 527; Conservative 197; Mismatches 449; Indels 139; Gaps 35;

```

QY 10 GLLALLPFGAA--STOYCTGTDMKRLRASPETHLDMRLHYOCGVQGNLEITYPT 67
DB 11 GLPFLARGSEVGNISQAVCPGLTNGLSVGDANOYQTLKYERCEVMKNLEIVLGH 70
QY 68 NASLSFLQDIOEVGYVLIHNOVROVPLQRLIYRGTLFEDNALAVLDNGPILNTT 127
DB 71 NADLSFLQWIREVTGYVLYMNEFTLPLNLRVYMGTVYDGKFAIVM-----LNYNT 125
QY 128 PVTGASPGGLREIQLSLTEILKGVLIQNPOLCYQDTILMKDIFKONQALTLIDTN 187
DB 126 ----NSSHALRQLRLTQLTEILSGVYIEKNDLCHMDTIDMDRIVDRD--AEIVKD 178
QY 188 RSRACHPCSPMKSGRSCWSESSEDCSLTRTVACAGC--ARCKGPLPTCCHEOCAGCTG 246
DB 179 NGRSCPCHEVCKG--RQWGPSEDQTLTKTCAPQNGHCGSPRNGCHDEOCGSGS 237
QY 247 PKISDCLACHENHSGICELHCPALVTYMTDFESPNRBEGYTFGASCYTAACPNTIST 306
DB 238 PDDTDFACHFNDISACVPCRPQPLVYNTKLTFOLEPNHTTYQYGVGCVASCPHNFV-V 296
QY 307 DVGSCITVCELANQVETABDGTQRCCKSPCARVCYGLMGYIKANSKF--IGITEL- 363
DB 297 DQTSVCVACPDPKMEVD--KXGLMKCEPCGGLCPKACEGIG-----SGSRFQTVSSNIDG 350
QY 364 FAGCKRTFGSLAFLPESFDPPASNTAPLQPELOVFELEITGYLYTSMWPSLPDLS 423
DB 351 FVNCTKLGNLDFLTGLNGDPWHKIPALDPEKLVNFRVREITTYLYIQSWPPHMFNS 410
QY 424 VFQNLQVIRGRILHNGAYS--LTLOGISWLGSLRSLREISLALIHNTLCEVHTYVW 482
DB 411 VFSNLTITIGRSLYNGFSLIMKKNLNTVSLSPRLKLSNRIRIYSANROLCTVHSLNW 470

```

```

QY 483 DQLFNPQALLHTA--NRPECEVGEIAGHOLCARGHCSWGPPTQCVNCSQFLRGQECV 541
DB 471 TKVLGRPEERLDIKHNRPRRDCVABEKVCDPLSSGCGCPGPGGGLSCRNRSRGVAV 530
QY 542 EECRVQLPREYVNAHCLPCHECOPONGSVYTCGPEADQCVACAHYKDPFCYARCP 601
DB 531 THCNPLNEPEPFAHEACEFSCHEPCPMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
QY 602 SGVXKDLVYMIWMPPEBEGACGCPINCTHSCVDLDDKCAPAEORA-----SPLTSIVA 657
DB 591 HGVVLG--AKGFIYKTPDVQNECRCHENCCTQSGKGPRLDQCLQTLVLIGKTLTALIV 648
QY 658 VGVLLV--VVLGVV---GILQYIKANSKF--GITLEPLTPSGAMPNQAQMLKET 710
DB 649 IAGLVVIMMIGGTFLYWRGRILQNKRAKMYLERGESIEPLDPS--EKANKVLARIKFT 707
QY 711 ELRKVYLGSGAFGVYKGIWIPGENVVKI PVAKIYVLRRENTSPKAKEILDEAYVWAGV 770
DB 708 ELRKVYLGSGVFTGKGVIPBGESEIKIPVCIKTIEDXSGQSFQAVTDHMLAIGSID 767
QY 771 SPYVRLIGLITSTVQLVQMLPFGCLLDHVENRGRIGSQDLLWCMQIAKMSYLED 830
DB 768 RAHIVRLTGLCPGSSSLQVLTQYPLBGLDHWQRGALCPQLLWNGVQIAKMYTLEE 827
QY 831 VLVYRDLAANVLYKSPNHVKITDPGLABLDIDETRYHADGKVPIKMMALSTLRR 890
DB 828 HGVYRNLAAKNVLLKSPSQVQVADFPVALPLPPDDQQLLYSEAKTPIKMALESIFK 887
QY 891 PTHQSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYVNC 950
DB 888 YTHQSDVSYGVTWELMTFGAEYAGLRLAEVPLLEKGERLAQOICTIDVYVNVAC 947
QY 951 WMTDECPRRRELVSERARAPQGFVYIYQNEUDLCPA---SPLDSTFRSLLEDDMG 1007
DB 948 WMTDINIRPTKELANETRYARAPRYLVYIKES--GGLIAPPEPHGLTKKLEVELE 1006
QY 1008 DLVDAEYLVPOQGFPCDPAPAGAGVYHRRHSSSTRSGGDLTLGLEP--SEEAAPSP 1066
DB 1007 PELDLDLLEBED-----NLATTTIGSLSLIPVGLNRPQSGSL 1047
QY 1067 LAPSEAGSDVFDGDLGMAAKGLQSLPTPD--PSPLQRYSEDTVLVLP-----SETDQY 1119
DB 1048 LSPSSGY--MPNQNLNLSGCEASVSSSERCPRVSLH-----BMPGCLASESSEGH 1100
QY 1120 VA-----PLTCSQPE-----YNOQPVRRPQPSPREGP-----SETDQY 1148
DB 1101 VTGSAELQEKVSMCRSRSRSRSPRPGDSAHYSQKSLTPTPLSPGLBEDVNGYV 1160
QY 1149 LPAAPFACATIERAKTISP--GRKGVV-----KDVAFGAVENPEYLTPOGCAAPQHP 1201
DB 1161 MDTHLAKGTSPSRGTLSSVGLSSVLGTBEDED-----EEVEYMRRRRHSP--PH 1211
QY 1202 PPAFPAFDNLVYWD-----QDPEFGAPSTFKGTPTAENBYL 1241
DB 1212 PRPSLEELGVEYWDVSDLSASLSTQSCPLRPVPRIMRAGTTPBEDYEM 1263

```

RESULT 10
 ERB3 RAT
 AC 062759; 062955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=852190;
 RX Heiliger N.J., Kim H.-H., Greaves C.H., Sterke S.L., Roland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85, 513 AND 565.
 RA Heiliger N.J., Roland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
 RT Expression of neuroligins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGLINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC33050.1; -
 DR HSP: P11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS0011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW transfease; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 644 662 POTENTIAL.
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 183 259 CYS-RICH.
 FT DOMAIN 707 964 PROTEIN KINASE.
 FT NP_BIND 713 721 ATP (BY SIMILARITY).
 FT BINDING 740 740 ATP (BY SIMILARITY).
 FT ACT_SITE 832 832 BY SIMILARITY.
 FT DISULFID 186 194 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402B8DFDIE CRC64;
 Query Match 34.2% Score 2316.5; DB 1; Length 1339;
 Best local Similarity 40.5%; Pred. No. 63e-117;
 Matches 520; Conservative 173; Mismatches 427; Indels 163; Gaps 36;
 QY 3 IAAALRWGLLLALLPPGAA---STQVCTGTDMLRLPASPETHLMDLRHLYOGCGVVOGN 59
 DB 7 LGVLC-----FLISLARSGEMGNSQVCGTGLNGLSVGDADNQYQTYKLYEKEVEVMGN 62
 QY 60 LELTYLPNNASLFDQDQVGVYLLAHNOVROPQLRLRYNGTQFEDNYALAVLDN 119
 DB 63 LEIVLTGNADISFLQWIREVTGYLVANMEFSVLPLPNLRVARGTVYQGKFAIFWM-- 120
 QY 120 GDPINNTPTVGTASPGGLRETLRLSTFELLKGVLLIGRNPLQCYQDTILMKDIFHKNNQL 179
 DB 121 ---LWYNT-----NSSHALRQLKFLQTLFELISGVYIERKDKLCHMDIIMRDIVRA--- 170
 QY 180 ALLTIDNRSAPCHPCSPCKGSGCSWSESSDCSLIRTYCAGGC-ARCGKPLPTDCHE 238
 DB 171 GAELIVKXNNGANCFPCHEVCKG-KCWGGPDDCQILTKTICAPQCNCRFCGPNMQCCHD 229
 QY 239 OCAAGCTGPKHSDCLACHFNHSGICELHCAVLATYNTDPEESWPNDEGRYTGAQCVTA 238
 DB 230 EAGAGCSGPQDIDCFACRRFVDSGACVPRCEPLVYNNKLTFLQLEPNPHTRYQYQGVCVAS 289
 QY 299 CPYNYLSTDVGSCTLVCPGLHNOETABDGTORCEKSPCARVYGL--GMQYIKANSKF 356
 DB 290 CPNHNV-VDQTFECVACPPDMEYD-KHGLMKCEPCGGLCPKACEGTGSSSRVGYVDSN 347
 QY 357 IGTILEPAGCKKIFGLSLAFPEFDDPASNTAPLPQLOVPEFTELETTGYLYISAMP 416
 DB 348 ID---GVNCTKTLGNIDFLITGLANDPMHKIPALDPEKLNFRYREITGYLNTQSWP 403
 QY 417 DSLPLSYFQMLQVIRGLIHNGAYS-LTQGLGISWLGSRSLRELOSGALITHNTHLC 475
 DB 404 PHMNFVSFSLUTTIIGRSLYNRFSLIKMKNLVTSGLRSLKEISAGVYISANQQLC 463
 QY 476 FVHYTPMQLFRNHQALLHTA-NRPDECVGSLACHQICARHGCMGPPTQVNCQSF 534
 DB 464 YHSHLMTNRLRLRGSERLDTKYDRPGLCEAGKVCDDPLCSSGGCGPQGLCLCRNY 523
 QY 535 LRQGCVECECVLQGLPREVYNARHCLPGEQCPQNGSVTCGSPAADQCVAAHYKDP 594
 DB 524 SREGVCTHNCFLQGEPRFVHAQCSCFPCLPMEGISTCGSGSDAARCAHRRDGE 583
 QY 595 FCVARCBGKVPDLSTYMPIMKFPDEBGAQCPBPINCTHSC--VDLDDKGGCPAQRASPLT 652

```

Db      584 HCVNSCPHCITG- AKGPIYKTPDQNECRPCHEHCITGCGPEIQDLCGAELMSKPH 641
Qy      653 STSAVVGI-LLVVLGVVF-----GILIQYIKANSKFI--GITEPLTPSGAMPNQMR 705
Db      642 LVIAVTGLAVITLMLIGSGFLVWRBRRIQNKRAMRVRIVERESTIEPLDPS-ERANKTLAR 700
Qy      706 ILKETELRKVKVLGGAGCTGYKGIWIDGENVKIPIVAIKYLRNTSPKANKETLDEAVY 755
Db      701 IFKETELRKVLGGVGFYKGIWIMEGESIKIPVICIKYIEKSGRGSFOATDTMFLA 760
Qy      766 MAGVSPYVSRLLGICLTSTVQLVQLMPYGCILDVHVENGRGLGSDLLNMCWQIKKM 825
Db      761 VGLSDHAHIVRLGLGCPSSQLDLYQYVPLISLDHKQKHETIGPQLLNWGVQIKKM 820
Qy      826 STLEDVRLVHRDLAARNLVKSPNHVKTITDPLGLRLDIDETEHADGKGVPIKMALES 885
Db      821 YLLEHSHVHRDLALRNLMKSPSQVQVADFGVADLLFPDDKQLHSEAKTPIKMALES 880
Qy      886 ILRRFTHQSVDVSGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLLPQPICITDVM 945
Db      881 IHFGKTHQSVDVSGVTVWELMTFGAPYAGLRLAETPDLEKGERLADPQICITDVM 940
Qy      946 IMVKCMWIDSECRPFRELVSEFSMARDPQRFVVIQNEIDLGPAAPLDSTFYRSLEDD 1005
Db      941 VWKCMWIDENIRPFKELANEFEMADPPRYLVIKAS-GPGRP--PAAEVSLTKE 997
Qy      1006 MGDVDAEEVLPQCGFCPPDPAQAGAMHHRSSSTSGGDDLTLGLEPSE----- 1060
Db      998 L-----QEALEPEL-----DLDLDLAEKEGLATS 1023
Qy      1061 -----EAPRSLAPSEG-----AGDVFDGLGWAAGKQSLPTH 1097
Db      1024 LGSALSLPTGTLTLPKRGQSLSPSSGVMNNSSLDGACLDASVLAGRGCFRPLSLH- 1082
Qy      1098 PSLPQRYSEDPTVLPSTDTGVY---APL-----TC-----SPOPE---YVNPQPV 1137
Db      1083 PIFGR-----PASESSEGHVGTSEAELOEKVYCRSRSRSPRPDRDSAYHSQRHS 1135
Qy      1138 RPOPPSPREP-----LPARPAGATLERAKTLSP-GNGVY-----KVFAF 1179
Db      1136 LLTPTPLSPGLEEDONGVMPDTHLRHASSREGITLSSVGLSTEDEED----- 1191
Qy      1180 GGAVERNPEYLTPOGGAAPQPHPP 1202
Db      1192 -----EEYVMNRKRGRSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EG 2.7.1.112) (Egfr)
DE (Garten receptor) (troped protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of allele cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RX Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baus J., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharjee D., Bolintiner S.,
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris J.M., Busam D.A., Butler H., Caden E., Center A., Chandra I.,
RA Chertis J.C., Dahlke S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Hattis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter D.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=DaekwanYeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Madsen S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RA "A Drosophila genomic sequence with homology to human epidermal
RT

```

```

RT growth factor receptor."
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RX ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/flb alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases."
RN Genetics 129:191-201(1991).
RP [9]
RX REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RL EGF receptor."
RL Cell 89:13-16(1997).
CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MARK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBryo, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -I- TISSUE SPECIFICITY: UNICITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTAGEN OF
CC TESTIS AND OVARY AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, AF052754; AAC08536.1; -.
DR EMBL, AF052753; AAC08536.1; JOINED.
DR EMBL, AF052754; AAC08535.1; -.
DR EMBL, AF052752; AAC08535.1; JOINED.
DR EMBL, K03054; AAA51462.1; -.
DR EMBL, K03417; AAA51460.1; -.
DR EMBL, K03416; AAA50985.1; -.
DR EMBL, K03418; AAA51461.1; -.
DR EMBL, AF109077; AAD26134.1; -.
DR EMBL, AF109078; AAD26132.1; -.
DR EMBL, AF109082; AAD26133.1; JOINED.
DR EMBL, AF109078; AAD26133.1; -.
DR EMBL, AF109084; AAD26133.1; JOINED.
DR EMBL, AF109079; AAD26130.1; -.
DR EMBL, AF109081; AAD26130.1; JOINED.
DR EMBL, AF109079; AAD26131.1; -.
DR EMBL, AF109083; AAD26131.1; JOINED.
DR EMBL, AF109080; AAD26135.1; -.
DR EMBL, AE003454; AAF46732.1; -.
DR EMBL, X02293; CAA26157.1; -.
DR EMBL, X78920; CAA55523.1; -.
DR EMBL, X78918; CAA55521.1; -.

```

```

DR EMBL, X78919; CAA55522.1; -.
DR PIR, A00640; GOFPE.
DR HSSP, P11362; IEGK.
DR Flybase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase_1.
DR Pfam; PF00757; Furin-like_1.
DR Pfam; PF01030; Recep_L_domain_2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SMO0219; FU; 7.
DR SMART; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.6%; Score 1935; DB 1; Length 1426;
Best Local Similarity 32.4%; Pred. No. 1.9e-96;
Matches 463; Conservative 183; Mismatches 428; Indels 354; Gaps 41;

QY 24 QVCTGTDMKLLPASPETHDMLRLHYQGCQVQGNLELYLP-NASLSFLDIDIOEVQ 82
DB 100 KICIGTKSRLSVPENKEHYENLRDYNCTYDGNLKLTLPLENIDLSFLLNIRVTV 159
QY 83 YVLNHNQVROVPORLRIVRGTLF-----EDNYALAVLDNGDPLNNTPTGASPG 137
DB 160 YILSHVYVKVPRPKQIIRGLRFLSLEBEKYLFLV-----TISKM 203
QY 138 RELQIRSLTEILKGVLIQRNPQCYDDTIWKDI FHKNNQALTLIDTRSPACHPSP 197
DB 204 YLLEIPDLRDVLNQGVGFHNNYNLCHMTIOWSBIIVNGTDAYNYDFTAPRECEPKCHE 263
QY 198 MCKGRGCMGSESSDCCSLTRTVGAGGCA--RCKGLPTDCCHEGCAAGCTGPHRSDCLAC 255
DB 264 SCTHG-CMGEGPKCKQKFSKLTGSPQCGAGRCYCPKREBCHLFCAGGCTGPTOKDCIAC 322
QY 256 LHFNHSIGCEIHLCPALVTYNTDTPESNPPEGRYTFGASCVTACPYNYLSTDVSGCTLVC 315
DB 323 KNFPEAVNSKECEPRMKRYNPTTVLLEINPEKGYAVGATCYKECP-CHLLRDQACVRS 381
QY 316 PLHNOEVTAEADGTQRCCKSKPCARVCYGLCMQYIKANSKFTIGTEL-----EPAGCK 369
DB 382 PQDKMDKQGE-----CVPNGPCPKTC-----PGTVYHAAGNDSFRNCTV 422
QY 370 IFGSLAFIPESFDG--DPASNTA-----PLQEPQLOVFLEIEITGYLYISAMPDPLPD 421
DB 423 IDGNRIIDQTFSGQDVYANYTMGPRIYPLDPRRREVFSTVEITGYLYISBETHQGFN 482
QY 422 LSVFQNLQVIRGLILHNGAY-SLTQGLGISWGLRSLRELQSGALIIHNHLCFVHY 480
DB 483 LSYFNLLETIRGRQIMESMFALAIIVSSYLSLEMRNLKQISSSVVIOHNRDLQVYSNI 542
QY 481 PWDQCFRPHQALLHTARPEDECVGSLGCHQICAGHCGMGPGPTCVVCSQFLGQEC 540
DB 543 RWRPAIQKPEQKWMNENLRADLEKRGITICSDQCNEDGCMGAGCTDCLTCKRNFNRGTC 602
QY 541 VEECRVLOGLPREYVNAHRLCPHRECQFQNGSVTCFGEPAQDCVACAHYKDPFCVARC 600

```

```

Db      603 IADCGYISNAK--FDNFTCKI CHPECK-----TCGAGADHCQCEVHVRDQGHVSEC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKNKYNDGVRCHATCGCTGPRDITIGACTTCNLAINNDATYKRCGLLMDKDPD- 713
Qy      609 SYMPIWKF--PDEBACQP-----CP1-----NCH----- 632
Db      714 GY--FMEXVHPQEQSLKPLAGRAVCRKCHPLCGLTNGYHEQVCSKCHYKRRECCT 721
Qy      633 -----SC-----VLDLDDKG----- 641
Db      712 ECPADHYTDEBQECGCRHPECNGCTGAGADCCSCNFKLFDANEIPRYNSTMPCST 831
Qy      642 -CPAEOR-----ASPLTS-----IVSAVGIILLVVLGVVFGI 673
Db      832 KCPLMRHVNQYTAIGPYCAASPSSSKITANLDVNMIFITIGAVLVPTICILCVTYI 891
Qy      674 LIQYIKANSKFIGIT-----ELPLPSGAMNQOMRIKETELRKXVVGSGAFCT 725
Db      892 CRQKQAKKEVYKMTALMSGCEDESEPLPSPNIGANLCKLRVQDAELRKGGVGMGAFGR 951
Qy      726 YKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAIVYVAGVSPYRLLIGICTLST 785
Db      952 YKGVVWPEGEENVKIPVAIKELKSTGAESESEFLREAYIMASEHVNLLKLAVCMSSQ 1011
Qy      766 VOLVQLMPYGCILDHVENRGRIGSDLLNWCQIAGKSYLEDVVLVHRDLAARNVLY 845
Db      1012 MMLITQLMPGLCLDYVNNNRDKIGSKALNWSQIAGKSYSEKELVHRDLAARNVLY 1071
Qy      846 KSPNHVKITDGLARLDDIDETEHADGKVPFKWMALESILRRFTHOSDVMVSYGVW 905
Db      1072 QTPSLVKITDGLAKLSSDSNEKYKAGGKMPKMLLECIRNRVFTSKSDVAFGTIIM 1131
Qy      906 ELMTFGAKPYDGIIPAREIPDLLENGERLPPPICTIVIMVYKCMNIDSECPFRRELY 965
Db      1132 ELITFGCRPHENIPAKDIPDLIEVGLKEQPEICSLDIYCTLLSCWMLDAMPTFKQLT 1191
Qy      966 SEFSMARDPQRFVYIQWEDLG--PASPBDSTFYSRLLED--DMWDLDVADAEVLPQO 1020
Db      1192 TVFAEFARDPGRYLAIPGDKTRIPA-----YTSQDEKDLRKLAFTDGSSEALAKPD 1244
Qy      1021 GFECDDPAPGAGGVNHRHRSSTRSGGDLTLGLEPSEEAR-----RSLAPSEGAG 1074
Db      1245 DYLQPKAPGS-----HRTDCT-----DEMPLKNYCKDPENKKSSTG 1283
Qy      1075 SDVPDG---DLQMGAKQLQSLPTHDSPLQRYSEDPVLEPSETDGYVAPLTCSPQPEY 1131
Db      1284 DDERSSAREVGVGNLR-----LDLPVDEDDYLMPTCGQFPNN 1321
Qy      1132 VNQPDVPRQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGAVENPEYL-- 1189
Db      1322 NNNNN-----NPNQNNMAAVGVAAGYM-----DLIGVPSVUNPEYLLN 1360
Qy      1190 --TPGGAAPOPH-----PPAFSP-AFDNLVYWD 1216
Db      1361 AQLTGVGESPIPTQITIGIPVWGPGTWEVVKVMPGSEPTSSDHEYND 1408

```

```

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP      SEQUENCE FROM N.A.
RX      MEDLINE=85228222; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-truncated EGF receptor.";
RL      Cell 41:719-726 (1985).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M10066; AAA8763.1; ALT_INIT.
CC      DR PIR: A06643; TVCHLV.
CC      DR PIR: B06643; TVEVLV.
CC      DR HSSP: P11362; 1FEK.
CC      DR InterPro: IPR0000719; Euk_Dkinase.
CC      DR InterPro: IPR001245; Tyr_Pkinase.
CC      DR Pfam: PF00069; Pkinase; 1.
CC      DR PRINTS: PR00109; TYRKINASE.
CC      DR ProDom: PD000001; Euk_Pkinase; 1.
CC      DR SMART: SMO0219; Tyrc; 1.
CC      DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC      DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
CC      DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC      KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      KW Glycoprotein; Phosphorylation.
CC      FT DOMAIN 132 399 PROTEIN KINASE.
CC      FT NP BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165 ATP (BY SIMILARITY).
CC      FT ACT SITE 257 257 BY SIMILARITY.
CC      SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

```

Query Match 24.6%; Score 1666.5; DB 1; Length 634;
Best Local Similarity 50.7%; Pred. No. 1.9e-82;
Matches 358; Conservative 81; Mismatches 138; Indels 129; Gaps 19;

```

```

Qy      567 CAHYKDPPEPCVAPCPGSKDPLSYMPIWKFPEDEBACQCPPICTHSCVLDLDDKCPAQ 646
Db      3 CAHFIDGHCYKAPGAVLGENDTL-VWKYADANAVCQLCHPNCTRGCCKPGDEGP--- 58
Qy      647 RASPLTISVAVV-GILVVVLGVVFGIIL--QYI---KANSKFIGITEL--PLTSGAM 698
Db      59 NGSKTPSIAAGVVGGLCLVAVVGLIGILYRKRHYRKTRLRLLQRELVLEPLTSGEA 118
Qy      699 PNOQMRILKETELRKXVVGSGAFGTIVYGIWIPQENYKIPVAIKVLRNTSPKANK 758
Db      119 PNOAHLRIKETELRKXVVGSGAFGTIVYGIWIPQENYKIPVAIKVLRNTSPKANK 178
Qy      759 ILDEAVYVAGVGSYVSRLLGICTSTVOVLTQMLPQCLLDHVENRGRIGSDLLNMC 818
Db      179 ILDEAVYVAGVGSYVSRLLGICTSTVOVLTQMLPQCLLDHVENRGRIGSDLLNMC 238
Qy      819 MQIAGMSYLEDVVLVHRDLAARNVLYKSPNHVKITDGLARLDDIDETEHADGKVP 878
Db      239 VQIAGMNVYLEERLVLVHRDLAARNVLYKTHQHKINDFGAKLGLGADKEVYHAGGKVP 298
Qy      879 KMALESILRRRFTHOSDVMVSYGVTWELMTFPAKPYDGIIPAREIPDLLEKGERLP 938

```



```

Db      299 KMALESLHRIYTHQSDVWSYGTVMELMTFGSKPFDGIPASISVLEKGRLLPOPI 358
Qy      939 CTIDVYIMVYKCMWIDSECRPRFRELVSSESRMARDQRFVVIQ-NEDLGPASPLDSTFY 997
Db      359 CTIDVYIMVYKCMWIDADSRKPRELLIAEFSKMARDPPRYLVIGDERMMLPSPTDSKFY 418
Qy      998 RSLLEDDDMGDLVAEELVYPOQGFPCDPAPAGAGVNHHRHSSSTRSGGDLTLGLEP 1057
Db      419 RFLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1058 SEEEAPRSPL-----APEGAGSDVFDGLDMGAKGLQSLPTHPSPLOQYSEDPVPL 1112
Db      450 -----SRTPLLSLATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPTGNR 496
Qy      1113 PSET--DGVAPLTCPOPEYVNOVDVPPPPSREGPLPAAPAGATLERAKTLSPGKN 1170
Db      497 LEESIDDFL-----PAPEYVNO--LMPKPS-----TAMVON 527
Qy      1171 GVVKDV-----AFGAVENPEYLTPOGGAAPQHPPEPAPAFDNLVYW 1215
Db      528 QYNNISLTAISKLPMSDRYQNSHSTAYDNPEYL-----NTNQSPLAKTYFESSPYW 579

Qy      1216 DQ-----DPEF-----RGAPSTFKCTPTAENPEYIGLDVP 1246
Db      580 IQSGNHQIINDMPDYQODFLPNETKPNGLKVPAPAEYILRVAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain E64).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RX NCBI Taxid=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RL gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debure B., Henry C., Benaisa M., Bisette G., Clavierie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC - MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC Thin SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR      EMBL: K02006; AAA42394.1; ALT_INIT.
DR      EMBL: K01216; AAA42400.1; -.
DR      PIR: A00644; TVYTH.
DR      HSSP: P11362; IEGK.
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR001245; Tyr_pkinase.
DR      Pfam: PF00069; pkinase.1.
DR      ProDom: PD000001; Euk_pkinase.1.
DR      SMART: SM00219; TyrcKc.1.
DR      PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR      PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
DR      TRANSFAC: TS50011; PROTEIN_KINASE_DOM.1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399
FT      NP_BIND 138 146 ATP (BY SIMILARITY).
FT      BINDING 165 165 ATP (BY SIMILARITY).
FT      ACT_SITE 257 257 BY SIMILARITY.
FT      CONFLICT 29 29 R -> W (IN REF. 2).
FT      CONFLICT 140 140 S -> F (IN REF. 2).
FT      CONFLICT 146 146 I -> V (IN REF. 2).
SQ      SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 23.9%; Score 1620; DB 1; Length 604;
Best Local Similarity 50.5%; Pred. No. 5,7e-80;
Matches 348; Conservative 78; Mismatches 131; Indels 132; Gaps 18;

Qy      587 CAHYKDPFPCVAPRGVGVPLSYMPMKFDEEGACQPCPINCCTHSQVLDJGCPAQ 646
Db      3 CAHFDGPHCVKACPAVGLGENDTL-VKRYADANAVQCLCPNCTRGCKPGLGCP 58
Qy      647 RASPLTSIVSAVV-GILLVVLGVVFGILI--QYI--KANSKFIGITEL--PLTPSGAM 698
Db      59 NGSKTPSIAGVAVGGLCLVYVVGIGILYRRRIYAKTLRRLCGRELVEPLTPSGEA 118
Qy      699 PNQAMRLKTELEARKVYVJGSAFGVYVGMIPQENKRIYAIKVLENTSPRANKE 758
Db      119 PNQALRLKTEERKRVKVGSGFGLYGLWLPBEKVKIPAIKLEATSPRANKE 178
Qy      759 ILDEAYVAGVGSVYSRLGICLTSTVQVLTQMLPFGCLLDHVRNRLGSGQDLNMC 818
Db      179 ILDEAYVAVSDNPHVCRKLGICLTSTVQVLTQMLPFGCLLDVIREKXNIGSQVYLNMC 238
Qy      819 MQIAKGSYLEDLVLRHDLAARVLYKSPNHVYKITDFGLARLLDIDETRYADGKVP 878
Db      239 VOIAKGMVYLERELVHRDLAARVLYKTPQHVYKITDFGLAKLIGADEKEXYHAGGVVP 298
Qy      879 KMALESLHRRFTHQSDVWSYGTVMELMTFGKRPYDGI PAEIPDLLEKGRLLPOPI 928
Db      299 KMALESLHRIYTHQSDVWSYGTVMELMTFGSKPFDGIPASEISSVLEKGRLLPOPI 358
Qy      939 CTIDVYIMVYKCMWIDSECRPRFRELVSSESRMARDQRFVVIQ-NEDLGPASPLDSTFY 997
Db      359 CTIDVYIMVYKCMWIDADSRKPRELLIAEFSKMARDPPRYLVIGDERMMLPSPTDSKFY 418
Qy      998 RSLLEDDDMGDLVAEELVYPOQGFPCDPAPAGAGVNHHRHSSSTRSGGDLTLGLEP 1057
Db      419 RFLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1058 SEEEAPRSPL-----APEGAGSDVFDGLDMGAKGLQSLPTHPSPLOQYSEDPVPL 1112
Db      450 -----SRTPLLSLATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPTGNR 496
Qy      1113 PSET--DGVAPLTCPOPEYVNOVDVPPPPSREGPLPAAPAGATLERAKTLSPGKN 1170
Db      497 LEESIDDFL-----PAPEYVNO--LMPKPSYAM----- 524
Qy      1171 GVVKDV-----GAVENPEYLTPOGGAAPQHPPEPAPAFDNL 1211
Db      525 -VQNDIVYFILTALISKLPMSDRYQNSHSTAYDNPEYL-----NTNQSPLAKTYFES 575
Qy      1212 LYWDQDPPERGAPSTFKCTAENPEY 1240

```

```

Db      576 SPYWIGSGNHQ-----INLNDPY 594
RESULT 14
ERBB_AVIEU ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273.
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13179; AAA42401.1; -.
DR PIR; A25231; TVPEVB.
DR HSSP; P11362; 1EGK.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA06B65D CRC64;
Query Match 22.8%; Score 1547; DB 1; Length 540;
Best Local Similarity 53.2%; Pred. NO. 4.2e-76;
Matches 329; Conservative 71; Mismatches 122; Indels 96; Gaps 16;

```

```

Db      179 ILDEAYVNASVDNHNVCRLGICLTSTVOLTOLMPYGCLLDIYREHKONIGSQYLLMNC 238
819 MOIAKMSYLEDVLRHDLAARNVLKSPNNHYKTFDGLARLLDIDETRYHADGKVP 878
239 VOIAKGMVLEERHMYRDLAARNVLKTOHQKIDFGAKQLGADDERHYHAGKVP 928
879 KMALESLIRRRFTHQSDWMSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGRLLPQPI 938
239 KMALESLIRHRYTHQSDWMSYGVTVLMTFGSKPYDGIIPASEISSVLEKGRLLPQPI 958
939 CTIVYVMYVCMWIDECRPFELVSEFSRMAPDQREVVIO-NEDLGAPASPLDSTFY 997
359 CTIDYVMYVCMWIDECRPFELVSEFSRMAPDQREVVIO-NEDLGAPASPLDSTFY 997
998 RSLLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGVWHRHSSSTRSGCDLTLGLEP 1057
419 RTLMEEDMEDIVDAEYLVHQGF-----NSPST----- 449
Db      1058 SEEARPSPL-----APSEAGSDVFVFGDLGMGAAGKQSLPHDPSPLORYSEDPVPL 1112
450 -----SRTPLLSSLSATSNNSATCIDRNQ-----H----- 476
OY      1113 PSETDGYVAPLTCSPQEEYVNVQDPVPPQPSPPGGLPAARPAAT-LERAKTLSPCKG 1171
Db      477 PVREDGL-----PAREYVNO-LPKKSTAMVQVINYISLTAISKLPIDSRQN- 527
OY      1172 VKQVFAFGAVENPEYL 1189
Db      528 -----SHSTAVDPEYL 539
RESULT 15
EGFR_CHICK ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Glavol D.,
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RT Mol. Cell. Biol. 8:1970-1978 (1988).
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@lsb-sib.ch.

```
CC -----
DR EMBL; M20386; AAA48760.1;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_PARTIAL.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_PARTIAL.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM_PARTIAL.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KV Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 1 30
FT DOMAIN 31 703
FT TRANSMEM 31 654
FT DOMAIN 655 667
FT DISULFID 668 >703
FT DISULFID 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 318 333
FT DISULFID 336 340
FT DISULFID 513 522
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 569 589
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match      22.3%; Score 1512; DB 1; Length 703;
Best Local Similarity 43.0%; Pred. No. 4,2e-74;
Matches 304; Conservative 113; Mismatches 254; Indels 36; Gaps 16;
```

```
QY 8 RWGLLALIPGAA-----STQVCTGIDMKLRLPASPEHLDMLRHLVYGCQVVGQNL 61
DB 13 RGAAYLVLLILGVALCSAVEEKVCOGINNKLTLQGHVEDHFTSLQRMYNKCEVALSNLE 72
QY 62 LTYLPTNASTFLQDIOEVGVVLAHNOVRCVPLQRIYRGTLQFEDNYALAVLNDG 121
DB 73 IYVEHNRDLTLKTIQEVAGVLLALNMVDVIFLENIQIRGNVLIDNSFALVLSNYH 132
QY 122 PLNNTPTVTGASPGSLRELQRLSTLEILKXGVLIQRPOLCYQDTILMKDI FHRNQGL 181
DB 133 -MNKTQ-----GIRELPMKRLSEIINGVKISNNPKLCNMDTVLMDIIDSRL-PL 182
QY 182 TLID-TNRSRACHPSPMCKSGSRCKGSESESDQSLTRTVCAAGCA-RCKGPLPTDCQHEQ 239
```

```
DB 183 TLVDFASNLSSCPKCHENCTEDHGMWAGNQCTLTIKVI CAQCCSGRCGRKVPSSDCCHNQ 242
QY 240 CAACCTGPKRSDCLACLFHNSGICELHCAFLTYNTDPFESNPNEGRYTGASCVTAC 299
DB 243 CAACCTGPRSDCLACKFRDDATCKQTCPLVLYNPTTYQMDVNEGKSGATVREG 302
QY 300 PNYLSTDVSGCTLVCPILHNOEYTAEDGTQRCCKSPCARVYGVGMQYIKANSKFIGI 359
DB 303 PHNYVTVTHSGCVASGNTDTYEV-EEGVYKCKCKCGGLCSKVCNGIGIGELKILS-INA 360
QY 360 TELE-FACCKKIFGSLAFLESPEDGDPASTATLQEQLOVEFTLEITGLYLSAWPOS 418
DB 361 TNDSFKKCKKINDVSLIPVAFPGAFTTLPDPRKLDVFRTVIEISGFLTIQAMPDN 420
QY 419 LPDVSFQNLQVIRGILHNGAVSLTLOGIGISMLGRLSRELGSGLALIHNTHLCFYH 478
DB 421 ATDLXAEENLEIIRKTRKQGYSLAVNLIKISLGRSLAKE:SDDDIAIMKRNLCYAD 480
QY 479 TVPDQLFRRNHQALLHTANRPEDECVGEGSLAQULCARGHCKGPGFTQCVCNSQFLRQ 538
DB 481 TMNRRSLFATQSQTKLIQRRNKKDCTADRHVCDPLCSVDGCMGPGPFHCFSCRFPSRQ 540
QY 539 ECVEGCVLQGLPREYVNAHCLPCHPEQCPONG---SYTCFGEADQCVACAHYKDPF 595
DB 541 ECVRQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKKHAFLDGR 600
QY 596 CVARCPGSKPEDLSPYPIWKFPPEBGAQCPPINCTHSVLDLQKCPAEQRPASPLTSIV 655
DB 601 CVKACPGVIGENLT-VMKYADANAVCQCHNCTRCGCKGPGLEGCP--NSKTFPSIA 656
QY 656 SAVY-GILLVVYLVGVFGILI--QYI---KANSKFIGITL--PLRP 694
DB 657 AGVVGILLCLVVGILGIGLYLRHHIVRKRLRLRLQERLVEPLRP 703
```

Search completed: July 22, 2003, 09:18:11
Job time : 16.5741 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.5261 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-653-675-12
Perfect score: 6776
Sequence: 1 MELALCRMGILLALIPGA.....TEKGTPTAENPEYLGDVDPV 1247

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6095	89.9	1259	6	018735 canis fami
2	3057	44.1	1209	11	090X70 ratu
3	3057	44.1	1210	11	09EP98 mus muscu
4	2646	39.0	1165	13	09YH40 xiphophor
5	2643.5	39.0	1137	13	09W656 gallu
6	2262	33.4	1328	13	P79754 fugu rubrip
7	2007.5	29.6	1433	5	09B1H9 anopheles g
8	1871	27.6	419	4	09UK79 mus sapien
9	1739	25.7	367	11	08R2X1 mus muscu
10	1697.5	25.1	412	4	08WYV0 homo sapien
11	1637	24.1	729	15	086712 avian trou-
12	1635	24.1	567	15	086714 avian trou-
13	1570.5	23.2	962	15	064895 avian eryth
14	1562	23.1	545	15	085468 avian eryth
15	1506.5	22.2	655	11	09WVF5 mus muscu
16	1490.5	22.0	643	11	09ERV6 mus muscu

17	1262	18.6	1193	5	09Y1X8 ephydaria f
18	1183.5	17.5	023821	5	023821 caenornabi
19	1144	16.9	026566	5	026566 schistosoma
20	1125	16.6	090836	13	090836 gallus gall
21	1001.5	14.8	478	11	09ES60 ratu
22	942.5	13.9	539	13	09PSH2 gallu
23	906	13.4	165	4	014256 homo sapien
24	887	13.1	176	11	0923V5 ratu
25	806.5	11.9	346	13	09W176 xiphophor
26	778	11.5	435	5	08SZM1 xiphophor
27	754.5	11.1	311	13	09V162 xenopus lae
28	743.5	11.0	1352	13	09V24 xenopus lae
29	734	10.8	331	4	09BUD7 homo sapien
30	727	10.7	1671	6	09N0V5 biophari
31	723	10.7	149	6	09B66 oryctolagus
32	698	10.3	1418	13	09J457 scophthalmu
33	688.5	10.2	1368	13	08UM85 paralicthy
34	676.5	10.0	1369	13	08UM86 paralicthy
35	672	9.9	1412	13	08UM84 paralicthy
36	668	9.9	1358	13	073798 xenopus lae
37	665.5	9.8	1472	5	09USAB bombyx mori
38	647.5	9.6	1418	13	08UM83 paralicthy
39	642	9.5	1245	13	09YGH8 scophthalmu
40	634	9.4	1371	11	09QVW4 ratu
41	615.5	9.1	2144	5	09VD94 drosophila
42	595	8.8	987	11	091YMO mus muscu
43	591.5	8.7	1091	4	09UMQ4 homo sapien
44	589	8.7	1036	4	007912 homo sapien
45	584	8.6	987	11	099MR2 mus muscu

ALIGNMENTS

RESULT 1
ID 018735 PRELIMINARY; PRT; 1259 AA.
AC 018735:
DT 01-JAN-1998 (TRENBLREL 05, Created)
DT 01-JAN-1998 (TRENBLREL 05, Last sequence update)
DT 01-JUN-2002 (TRENBLREL 21, Last annotation update)
DE Erdb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBT_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "CDNA cloning of erdb-2 from canine mammary gland."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008451; BAA23127.1; -
DR HSR; P11362; IFC.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR InterPro: IPR004019; Furin-like.
DR Pfam: PF00757; Pkinase_1.
DR Pfam: PF00069; Pkinase_1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART; SMO0261; FU; 3.
DR SMART; SMO0219; TYRC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match		89.8%; Score 6095; DB 6; Length 1259;
Best Local Similarity		89.5%; Pred. No. 0;
Matches 1128; Conservative		46; Mismatches 72; Indels 14; Gaps 5;
Qy	1	MELALCRWGLLLALPPGAASVQVCTGDMKRLPASPETHLDMLRHLYOGQVQVQNTL 60
Db	1	MELAMCRWGLLLALLPSGAAGVQVCTGDMKRLPASPETHLDMLRHLYOGQVQVQNTL 60
Qy	61	ELTYLPTNASLSFLQDIQEVQGYLLAHNOVROVPLORLRYVGTQLFEDNYALAVLDNG 120
Db	61	ELTYLPANASLSFLQDIQEVQGYLLAHNSQVROPLORLRYVGTQLFEDNYALAVLDNG 120
Qy	121	DPLNNTPTVGTASPGRLQELQSLTEILKGYLLQRPQLCYQDTILMKDIFHNQOLA 180
Db	121	DPLEGGIPAPGAAGGRLQELQSLTEILKGYLLQRPQLCYQDTILMKDIFHNQOLA 180
Qy	181	LTLLIDNRSRACHCSPCKGSRCKWGESSEDCOSLRTRVAGGACRCKPLPTDCHEOC 240
Db	181	LTLLIDNRSACHCSPCKGSRCKWGESSEDCOSLRTRVAGGACRCKPLPTDCHEOC 240
Qy	241	AAGCTGPRGSDCLALPHNSGICELHCPALVTYNTDTEFSEMPNPRGRYTFGASCVTACP 300
Db	241	AAGCTGPRGSDCLALPHNSGICELHCPALVTYNTDTEFSEMPNPRGRYTFGASCVTACP 300
Qy	301	YNYLSTDVSGCTLVCPILHNOVTAEDTQCEKCKSRCAVCGIGLGMQYIKANSKFIQIT 360
Db	301	YNYLSTDVSGCTLVCPILHNOVTAEDTQCEKCKSRCAVCGIGLGMQYIKANSKFIQIT 360
Qy	361	ELFAPGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVETLEETGYLTSAMPDSL 420
Db	361	IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVETLEETGYLTSAMPDSL 420
Qy	421	DLVFNQLQVIRGILHNGAVSLTLQGLISWGLKSLRPLSGGLLIHNTHLCEVHTV 480
Db	421	NLSVFONLRVIRGVLHDGAVSLTLQGLISWGLKSLRPLSGGLLIHNTHLCEVHTV 480
Qy	481	PMQDLFRPHQALHTANRPECDEVCSEGLACHOLCARGHCWGGPQCVCNCSOFLRGQCC 540
Db	481	PMQDLFRPHQALHTANRPECDEVCSEGLACHOLCARGHCWGGPQCVCNCSOFLRGQCC 540
Qy	541	VEECRVLOGLPREVYANRHCIPCHPECCOPNGSVTCGEPADOCVACAHYKDPFCVAC 600
Db	540	VEECRVLOGLPREVYANRHCIPCHPECCOPNGSVTCGEPADOCVACAHYKDPFCVAC 600
Qy	601	PSGVKPDLSYMPIKFPDEBEGACQPCINCTHSCVDLDGCAEQASAPLSIVAAVG 660
Db	600	PSGVKPDLSYMPIKFPDEBEGACQPCINCTHSCVDLDGCAEQASAPLSIVAAVG 660
Qy	661	ILLVVLGVVFGILI---QYIKANS--KEIGITEL--PLTPSGAMPNOAQRILKETEL 712
Db	660	ILLVVLGVVFGILIKRRQKIRKYTRRLQETELVEP.LTPSGAMPNOAQRILKETEL 712
Qy	713	RKVYVLSGAGTGYKIKWIPDGNVYKIPVAILKYLENTSPKANKELIDBAYMAVAGVSP 772
Db	720	RKVYVLSGAGTGYKIKWIPDGNVYKIPVAILKYLENTSPKANKELIDBAYMAVAGVSP 772
Qy	773	YVSRLLGICLSTVQLTOLMPYCLLDHVENRGRIGSODLNMCKQIAKGSYLEDR 832
Db	780	YVSRLLGICLSTVQLTOLMPYCLLDHVENRGRIGSODLNMCKQIAKGSYLEDR 832
Qy	833	LVHRDLAARNVLVSPNHVKITDGLARLLDIDETRYHADGCVPIKMALESILRRRT 892
Db	840	LVHRDLAARNVLVSPNHVKITDGLARLLDIDETRYHADGCVPIKMALESILRRRT 892
Qy	893	HOSDVMASGYVMEIMTFGAPYDGIAREIPDLLEKGEELPQPICTIDVYIMVCKM 952
Db	900	HOSDVMASGYVMEIMTFGAPYDGIAREIPDLLEKGEELPQPICTIDVYIMVCKM 952
Qy	953	IDSECRPREFELVSEFSRMADPQRFVIVIQNEDIGPASPIDSTFFYSLLEDMDGLVDA 1012
Db	960	IDSECRPREFELVSEFSRMADPQRFVIVIQNEDIGPASPIDSTFFYSLLEDMDGLVDA 1012
Qy	1013	EFLVLPQGGFFCPRPAGAGGVYHHRSSSTRSGGDLTLGLPEBEEAPSPILAPSEG 1072

Db	1020	EFLVLPQGGFFCPRPAGAGTARHRRSSSTRNGGSLTLGLPSEEEPPKSPILAPSEG 1079
Qy	1073	AGSDVPBGDLGMAAGLQSLPHNDSPLOQVSEDPVTLVSETOGYVAPLTCSPQPEV 1132
Db	1080	AGSDVPBGDLGMAAGLQSLPHNDSPLOQVSEDPVTLVSETOGYVAPLTCSPQPEV 1132
Qy	1133	NQPVPRQPSPEGLPAPAPGATLER----AKTLSPKXGVKGVDFARQAVENPE 1187
Db	1140	NQPVPRQPSPEGLPAPAPGATLER----AKTLSPKXGVKGVDFARQAVENPE 1187
Qy	1188	YLPFGGAAPQHPHPAFSPAFNLVYMDQDPERGAPSTGTGPTAENPEYGLDVPV 1247
Db	1200	YLPFGGAAPQHPHPAFSPAFNLVYMDQDPERGAPSTGTGPTAENPEYGLDVPV 1259
RESULT 2		
Qy	090X70	PRELIMINARY; PRT; 1209 AA.
Db	090X70	PRELIMINARY; PRT; 1209 AA.
AC	090X70	01-MAY-2000 (TREMBlrel. 13, Created)
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OK	NCBI_Taxid=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=9025888; PubMed=2342466;	
RA	Barn H.S.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;	
RA	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K.; Dawson T.L.; Barn H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -	
DR	HSP; P1362; JRG.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_Pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_Pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; Pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_Pkinase; 1.	
DR	SMART; SM00261; Fv; 3.	
DR	SMART; SM00219; Tyrc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS0011; PROTEIN KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SO	SEQUENCE 1209 AA; 134891 MW; 96FE7FECCLB7773 CRC64;	
Query Match		45.1%; Score 3057; DB 11; Length 1209;
Best Local Similarity		49.2%; Pred. No. 1.2e-220;
Matches 625; Conservative		171; Mismatches 359; Indels 115; Gaps 29;
Qy	11	LLALLPAGAA--STVCGTDMKRLPASPETHLDMLRHLYOGQVQVQNTLTYLPTN 68